

3

STIC-Biotech/ChemLib

145029

mg

From: Li, Bao-Qun
Sent: Monday, February 14, 2005 8:43 AM
To: STIC-Biotech/ChemLib

PLEASE DO THE SEQUENCE HOMOLOGY AND INTERFERENCE OF SEQ ID NO; 20 IN APPLICATION SN.
09,664,363. THANKS.

Bao Qun Li M.D
TC 1600
Art Unit 1648
Tel. 517-272-0904
REM, 3C18
Rm. 3D24

RECEIVED
FEB 16 2005
STIC

STAFF USE ONLY

Searcher: Arnold
Searcher Phone: 2-2532
Date Searcher Picked up: 2/12/05
Date Completed: 2/12/05
Searcher Prep/Rev. Time: 2/12/05
Online Time: _____

Type of Search

NA Sequence: # 1
AA Sequence: # _____
Structure: # _____
Bibliographic: _____
Litigation: _____
Patent Family: _____
Other: _____

Vendors and cost where applicable

STN: _____
DIALOG: _____
QUESTEL/ORBIT: _____
LEXIS/NEXIS: _____
SEQUENCE SYSTEM: _____
WWW/Internet: _____
Other(Specify): _____

This Page Blank (u.s.p.),



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 145029

TO: Bao-Qun Li
Location: rem/3a24/3c18
Art Unit: 1648
Monday, February 28, 2005

Case Serial Number: 09/664363

From: Deirdre Arnold
Location: Biotech-Chem Library
REM 1A64
Phone: 571-272-2532

Deirdre.Arnold@uspto.gov

Search Notes

Please feel free to contact me if you have any questions or would like to amend the search.

Thank you for using STIC services.

Regards,
Deirdre Arnold

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2005, 02:17:34 ; Search time 8554 Seconds
(without alignments)
11572.832 Million cell updates/sec

Title: US-09-664-363-20
Perfect score: 2043
Sequence: 1 TGGGAGGCGCTTACACAG.....TGTGGCGCATGAGATGGGC 2043

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapept 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_to:*
11: gb_scs:*
12: gb_ey:*
13: gb_un:*
14: gb_vl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Query Length | ID | Description |
|------------|--------|-------------|--------------|----|-----------------------------|
| 1 | 2043 | 100.0 | 2043 | 6 | A28153 PT-NANBH mr |
| 2 | 2043 | 100.0 | 2043 | 6 | A32201 NANBH PT P |
| 3 | 2043 | 100.0 | 2043 | 6 | AR144049 Sequence |
| 4 | 2043 | 100.0 | 3750 | 6 | A28157 PT-NANBH mr |
| 5 | 2043 | 100.0 | 3750 | 6 | AR144051 Sequence |
| 6 | 2035 | 99.6 | 7065 | 6 | A76577 Sequence 9 |
| 7 | 2022.2 | 99.0 | 3750 | 6 | A32203 NANBH PT P |
| 8 | 1804.6 | 88.3 | 9410 | 14 | HPCKR152 D50485 Hepatitis C |
| 9 | 1802.6 | 88.2 | 9616 | 14 | AB049088 Hepatitis C |
| 10 | 1801.4 | 88.2 | 9587 | 14 | AF333324 Hepatitis C |
| 11 | 1798.2 | 88.0 | 9379 | 14 | AF165060 Hepatitis C |
| 12 | 1796.6 | 87.9 | 7863 | 6 | AR005087 Sequence |
| 13 | 1796.6 | 87.9 | 7863 | 6 | AR064526 Sequence |
| 14 | 1796.6 | 87.9 | 7863 | 6 | AR091919 Sequence |
| 15 | 1796.6 | 87.9 | 7863 | 6 | AR145568 Sequence |
| 16 | 1796.6 | 87.9 | 7917 | 6 | AR005085 Sequence |
| 17 | 1796.6 | 87.9 | 7917 | 6 | AR064524 Sequence |
| 18 | 1796.6 | 87.9 | 7917 | 6 | AR091917 Sequence |
| 19 | 1796.6 | 87.9 | 7917 | 6 | AR145566 Sequence |

| | | | | | |
|----|--------|------|------|----|-----------------------------|
| 20 | 1796.6 | 87.9 | 9030 | 6 | AR005076 Sequence |
| 21 | 1796.6 | 87.9 | 9030 | 6 | AR064515 Sequence |
| 22 | 1796.6 | 87.9 | 9030 | 6 | AR091908 Sequence |
| 23 | 1796.6 | 87.9 | 9030 | 6 | AR145557 Sequence |
| 24 | 1796.6 | 87.9 | 9416 | 6 | AR005070 Sequence |
| 25 | 1796.6 | 87.9 | 9416 | 6 | AR064509 Sequence |
| 26 | 1796.6 | 87.9 | 9416 | 6 | AR091902 Sequence |
| 27 | 1796.6 | 87.9 | 9416 | 6 | AR145551 Sequence |
| 28 | 1796.6 | 87.9 | 9416 | 6 | E08315 cDNA encodi |
| 29 | 1796.6 | 87.9 | 9416 | 14 | HPCKR152 D50485 Hepatitis C |
| 30 | 1795 | 87.9 | 9374 | 14 | AF207753 Hepatitis C |
| 31 | 1795 | 87.9 | 9386 | 14 | AF165055 Hepatitis C |
| 32 | 1795 | 87.9 | 9416 | 6 | E10035 cDNA of non |
| 33 | 1795 | 87.9 | 9573 | 14 | AB049090 Hepatitis C |
| 34 | 1793.4 | 87.8 | 9379 | 14 | AF165059 Hepatitis C |
| 35 | 1793.4 | 87.8 | 9379 | 14 | AF165059 Hepatitis C |
| 36 | 1793.4 | 87.8 | 9386 | 14 | AF165056 Hepatitis C |
| 37 | 1788.6 | 87.5 | 9379 | 14 | AF207767 Hepatitis C |
| 38 | 1788.6 | 87.5 | 9410 | 14 | HPCKR152 D50481 Hepatitis C |
| 39 | 1787 | 87.5 | 7989 | 6 | AR406043 Sequence |
| 40 | 1787 | 87.5 | 7989 | 6 | AX036255 Sequence |
| 41 | 1787 | 87.5 | 7989 | 12 | SSE242652 Hepatitis C |
| 42 | 1787 | 87.5 | 8001 | 6 | AR406042 Sequence |
| 43 | 1787 | 87.5 | 8001 | 6 | AX036254 Sequence |
| 44 | 1787 | 87.5 | 8001 | 12 | SSE242654 Hepatitis C |
| 45 | 1787 | 87.5 | 8637 | 6 | AR406041 Sequence |

ALIGNMENTS

RESULT 1
A28153
LOCUS PT-NANBH mRNA fragment from patent GB229245.
ACCESSION A28153
VERSION A28153.1 GI:1248636
KEYWORDS
SOURCE
ORGANISM
synthetic construct
other sequences, artificial sequences.
REFERENCE
1 (bases 1 to 2043)
AUTHORS
TITLE
JOURNAL
FEATURES
Source
Post-translational non-A non-B hepatitis viral polypeptides
Patent: GB 2239245-A 20 26-JUN-1991;
location/Qualifiers
1..2043
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/clone="156/192"
<1..>2043
/codon_start=1
/transl_table=1
/product="PT-NANBH polypeptide"
/protein_id="CAA01930.1"
/db_xref="gi:1248637"

CDS

ORIGIN
Query Match 100.0%; Score 2043; DB 6; Length 2043;
Best Local Similarity 100.0%; Pred. No. 0;
TAAATKRIATAGSPPLASSASQLSAPSSKATYITQNDPFDADLIENLMMRMWG

DEFINITION NANBHV PT polyprotein fragment.
 VERSION A32201
 KEYWORDS A32201.1 GI:1926560
 SOURCE Non-A, non-B hepatitis virus
 ORGANISM Non-A, non-B hepatitis virus
 REFERENCE Viruses; unclassified viruses.
 AUTHORS 1 (bases 1 to 2043)
 JOURNAL
 FEATURES
 Patent: FR 2655990-A 20 21-JUN-1991;
 Location/Qualifiers
 1..2043
 /organism="Non-A, non-B hepatitis virus"
 /mol_type="unassigned DNA"
 /db_xref="taxon:12440"
 /clone="156/192"
 <1..>2043
 /codon_start=1
 /product="partial NANBH-PT polyprotein"
 /protein_id="CAA02093.1"
 /db_xref="GI:1926561"
 /translation="MEGVPTGLTHYDAHRLSOTKQAGNPVLAIVQATVCARQAP
 PSHDMKCLIRLKPPLRGPPLILRLGAVQNEVTLTPITPFIACNSADLEVTST
 WLVGVLAALAAVCLTGSVVIIGRIILSGPAIVPREVLVQEFDEMECSHLPI
 IEQWQLAEQFQKALGLQTAKEAAPVRESKRALLETFMKHMNFISQIOLY
 AGSLPNSIPALISLMAFTASVPLTOSTLLNLIGMVAQOLAPESASAIFYAG
 IAGAAGSIGIKVLVDILAGYAGAGALVAFKMGSEMPSTEDLVNLPAISPA
 LVGVYCAAILLRHYPREGAVQNMNRIAFASRGNHSPHYVESDAAYTQIS
 DLTQQLRLHOMINEDSTPCSGMLVDVDMITCTVADKTLVQSLRLPLGVP
 FFSCQRYKGVNRGDGIMQTTSCGAQITGHVKNQSMRLVDEKTCNNMHGTFINAY
 TTGCTSPAPNYSRALMRVAEVEYVGVDFHYVISMSTDNVKPCQVAPPEFT
 EVDGVRHRYAPACKPLREEVTFQVGLNVOYVGSQLPCEPEPDVAVLISMLTPDSHI
 TAFTRRLARSPSPSIASSSSQSLAPSSKATYITQNDFPADLIENLIRHMG"

CDS

Query Match 100.0%; Score 2043; DB 6; Length 2043;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2043; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ORIGIN

1 TGGGAGGGGCTCTTCAACAGGCTCAACCCAGTGGATGCCACTTCTGTCACCAACAAAG 60
 Db 1 TGGGAGGGGCTCTTCAACAGGCTCAACCCAGTGGATGCCACTTCTGTCACCAACAAAG 60
 QY 61 CAGGAGAGAGCAACTTCCCTACCTGAGGGGTACAGGCTTCTGTCAGGCTAAAGCTTACT 120
 Db 61 CAGGAGAGAGCAACTTCCCTACCTGAGGGGTACAGGCTTCTGTCAGGCTAAAGCTTACT 120
 QY 121 CAGGAGAGAGCAACTTCCCTACCTGAGGGGTACAGGCTTCTGTCAGGCTAAAGCTTACT 180
 Db 121 CAGGAGAGAGCAACTTCCCTACCTGAGGGGTACAGGCTTCTGTCAGGCTAAAGCTTACT 180
 QY 181 CTGCGCGGGGCCAACACCTTGTCTGTATAGGCTGGAGCGCTCCAAAACAGAGTCAACCTTC 240
 Db 181 CTGCGCGGGGCCAACACCTTGTCTGTATAGGCTGGAGCGCTCCAAAACAGAGTCAACCTTC 240
 QY 241 ACACACCCCATTAACAAATTTCATATGAGATGATGTGACGCGACCTGAGAGTGTCAAG 300
 Db 241 ACACACCCCATTAACAAATTTCATATGAGATGATGTGACGCGACCTGAGAGTGTCAAG 300
 QY 301 AGCACCTGGGTGCTGGTGGCGGGGTCTTGCAGCTCTGGCTGCGATATGTTGTTCCGAC 360
 Db 301 AGCACCTGGGTGCTGGTGGCGGGGTCTTGCAGCTCTGGCTGCGATATGTTGTTCCGAC 360
 QY 361 GGCACGCTGTGATTTGTGGTAGATCATCTTGTCCGGCGGGCTATTGTTCCCGAC 420
 Db 361 GGCACGCTGTGATTTGTGGTAGATCATCTTGTCCGGCGGGCTATTGTTCCCGAC 420
 QY 421 AGGGAAGTCTCTACAGAGATTGATGATGAGAAAGTGGCGGTGCTCCCTTAC 480
 Db 421 AGGGAAGTCTCTACAGAGATTGATGATGAGAAAGTGGCGGTGCTCCCTTAC 480
 QY 481 ATCGAGAGGAATCAGCTCCCGAGCAGTTCAAGCAAAAAGCGCTCGGTTGCTGAC 540

Db 481 ATCGAGAGGAATCAGCTCCCGAGCAGTTCAAGCAAAAAGCGCTCGGTTGCTGAC 540
 QY 541 ACAGCCCAACCAAGCAGGAGCGCTGTCCCGTGTGAGTCCAAAGTGGAGCCCTT 600
 Db 541 ACAGCCCAACCAAGCAGGAGCGCTGTCCCGTGTGAGTCCAAAGTGGAGCCCTT 600
 QY 601 GAGACCTTCTGGGCGAAACACATGTGAACTTCAATCAGCGGATACAGTACTTACAGGC 660
 Db 601 GAGACCTTCTGGGCGAAACACATGTGAACTTCAATCAGCGGATACAGTACTTACAGGC 660
 QY 661 TTGTCACTCTGCTCGGGAATCCCGGATGTGATCACTGATGCGCTTCAACCTCTGTC 720
 Db 661 TTGTCACTCTGCTCGGGAATCCCGGATGTGATCACTGATGCGCTTCAACCTCTGTC 720
 QY 721 ACTAGCCGCTCACCACCAATCTACCCCTCGTTAATCACTCTGAGGAGATGGATGCC 780
 Db 721 ACTAGCCGCTCACCACCAATCTACCCCTCGTTAATCACTCTGAGGAGATGGATGCC 780
 QY 781 GCCCAACTGCTCCCCAGTGTCTCTTCAAGTGTGAGGCGCGCATTTGCTGTGCG 840
 Db 781 GCCCAACTGCTCCCCAGTGTCTCTTCAAGTGTGAGGCGCGCATTTGCTGTGCG 840
 QY 841 GCTGTTGGCAGATAGGCTTGGGAAAGTGTGAGACATTTGGCGGCTATGAGAG 900
 Db 841 GCTGTTGGCAGATAGGCTTGGGAAAGTGTGAGACATTTGGCGGCTATGAGAG 900
 QY 901 GGAAGTGGAGAGGCGCTGTGGCTTTAAGTCATGAGGCGGAATGCCCTCCACCGAG 960
 Db 901 GGAAGTGGAGAGGCGCTGTGGCTTTAAGTCATGAGGCGGAATGCCCTCCACCGAG 960
 QY 961 GACCTGTTAACTTACTCCCTGCAATCCCTCTCTCTGCTGCTGCTGCTGCTGCTGCTG 1020
 Db 961 GACCTGTTAACTTACTCCCTGCAATCCCTCTCTCTCTGCTGCTGCTGCTGCTGCTGCTG 1020
 QY 1021 TGGCAGAGGATATCTGCTGCGGACCTGAGGAGAGGAGGCTGTGAGTGAAC 1080
 Db 1021 TGGCAGAGGATATCTGCTGCGGACCTGAGGAGAGGAGGCTGTGAGTGAAC 1080
 QY 1081 CGGCTGATAGCGTGTGCTGCGGGGTAAACATGTTTCCCGACGACATATGTCAGAG 1140
 Db 1081 CGGCTGATAGCGTGTGCTGCGGGGTAAACATGTTTCCCGACGACATATGTCAGAG 1140
 QY 1141 AGCGAGCGCGAGCAGTGTCACTAGATCCCTCTCCGACTTATCATCAACCACTGTTG 1200
 Db 1141 AGCGAGCGCGAGCAGTGTCACTAGATCCCTCTCCGACTTATCATCAACCACTGTTG 1200
 QY 1201 AAGAGCTTCCACAGTGAATTAAAGAGATGTCTCACGCTGTCCGGCTGTGGCTA 1260
 Db 1201 AAGAGCTTCCACAGTGAATTAAAGAGATGTCTCACGCTGTCCGGCTGTGGCTA 1260
 QY 1261 AGGATGTTTGGGACTGTGATATGCAAGTTTGGCTGAACCTTGGCTCCAGTCC 1320
 Db 1261 AGGATGTTTGGGACTGTGATATGCAAGTTTGGCTGAACCTTGGCTCCAGTCC 1320
 QY 1321 AAGCTCTGCGCGATTAACGGGAGTCCCTTTTCTCATGCGCAAGTGGATCAAGGGG 1380
 Db 1321 AAGCTCTGCGCGATTAACGGGAGTCCCTTTTCTCATGCGCAAGTGGATCAAGGGG 1380
 QY 1381 GTCTGGCGGGAGACGGGATCATGAGACCACTGCTCATGTGGAGCACAGTACCGGA 1440
 Db 1381 GTCTGGCGGGAGACGGGATCATGAGACCACTGCTCATGTGGAGCACAGTACCGGA 1440
 QY 1441 CATGTCAAAAACGGTTCATAGAGATGTTGGGCTTAAGACCTGTATGTAATGTGGCAT 1500
 Db 1441 CATGTCAAAAACGGTTCATAGAGATGTTGGGCTTAAGACCTGTATGTAATGTGGCAT 1500
 QY 1501 GGAACATTCCTCATAGCATACACAGCGGAGCCCTGACAGCCCTCCAGCGCAAC 1560
 Db 1501 GGAACATTCCTCATAGCATACACAGCGGAGCCCTGACAGCCCTCCAGCGCAAC 1560
 QY 1561 TATTCCAGGCGCTGTGGCGGGTGTCTGTGAGGATACGTGAGTAAACGGGGTGGGG 1620
 Db 1561 TATTCCAGGCGCTGTGGCGGGTGTCTGTGAGGATACGTGAGTAAACGGGGTGGGG 1620

QY 1621 GATTTCACCTAGTGAGCAGATGACCACTGACAAAGTAAATGCCCCGTGCCAGGTTCCA 1680
DB 1621 GATTTCACCTAGTGAGCAGATGACCACTGACAAAGTAAATGCCCCGTGCCAGGTTCCA 1680
QY 1681 GCCCCGGAATTTTTCACAGAAATGAGATGGGGTGGCGCTGCACAGATGAGCTTCGGCGTGC 1740
DB 1681 GCCCCGGAATTTTTCACAGAAATGAGATGGGGTGGCGCTGCACAGATGAGCTTCGGCGTGC 1740
QY 1741 AAACCTCTCCATCGGAGAGAGGTGACATTTCCAGATCGGGCTCAACCAATACCTGGTGGG 1800
DB 1741 AAACCTCTCTACGAGAGAGAGGTGACATTTCCAGATCGGGCTCAACCAATACCTGGTGGG 1800
QY 1801 TCGAGCTTCCATGCGAGGCCGACCGAATGATGACAGTGTCACTTCCATGCTCACCGAC 1860
DB 1801 TCGAGCTTCCATGCGAGGCCGACCGAATGATGACAGTGTCACTTCCATGCTCACCGAC 1860
QY 1861 CCGTCCCATCATACAGCAGAGACGGCTAAGCGGACAGCTGGCCAGGGGGTCTCCCTCC 1920
DB 1861 CCGTCCCATCATACAGCAGAGACGGCTAAGCGGACAGCTGGCCAGGGGGTCTCCCTCC 1920
QY 1921 TTGGCAGCTTCTTCCAGCTAGCCAGTTGTCTGGCCCTTCTCTCGAAGGCGACATACATTACC 1980
DB 1921 TTGGCAGCTTCTTCCAGCTAGCCAGTTGTGTGGCCCTTCTCTCGAAGGCGACATACATTACC 1980
QY 1981 CAAAATGACTTCCAGACGCTGACCTCATCGAGGCCAACCTCTGTGGCGGATGAGATG 2040
DB 1981 CAAAATGACTTCCAGACGCTGACCTCATCGAGGCCAACCTCTGTGGCGGATGAGATG 2040
QY 2041 GGC 2043
DB 2041 GGC 2043

RESULT 3
AR144049
LOCUS AR144049 2043 bp DNA linear PAT 08-AUG-2001
DEFINITION Sequence 20 from patent US 6210675.
ACCESSION AR144049
VERSION AR144049.1 GI:15105916
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2043)
AUTHORS Highfield, P.Edmund., Rodgers, B.,Colin., Tedder, R.,Seton. and Bairdara, J.,Anthony.,James.
PT-NANB hepatitis B polyepitides
JOURNAL Patent: US 6210675-A 20 03-APR-2001;
FEATURES Location/Qualifiers
source 1..2043
/organism="unknown"
/mol_type="unasigned DNA"

ORIGIN
Query Match 100.0%; Score 2043; DB 6; Length 2043;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2043; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGGAGGGGCGTCTTTCACAGGCTTACCCACGTGATGCCACTTCTGCTCCAAACAAG 60
DB 1 TGGGAGGGGCGTCTTTCACAGGCTTACCCACGTGATGCCACTTCTGCTCCAAACAAG 60
QY 61 CAGGAGAGAGACAATTTCCCTTACCTGGTGGGTGACAGGCTAATCTGTGGCTTGGGCG 120
DB 61 CAGGAGAGAGACAATTTCCCTTACCTGGTGGGTGACAGGCTAATCTGTGGCTTGGGCG 120
QY 61 CAGGAGAGAGACAATTTCCCTTACCTGGTGGGTGACAGGCTAATCTGTGGCTTGGGCG 120
DB 61 CAGGAGAGAGACAATTTCCCTTACCTGGTGGGTGACAGGCTAATCTGTGGCTTGGGCG 120
QY 121 CAGGCCCCCAGCTTCATCATGGAGTCAAAATGTGGAATGTCTCATACGCTTAAAGCTTACT 180
DB 121 CAGGCCCCCAGCTTCATCATGGAGTCAAAATGTGGAATGTCTCATACGCTTAAAGCTTACT 180
QY 181 CTGGCGGGGCCAACAACCTTGTGTATAGGCTGGGAGCCGTCGAAAACGAGGTCAACCTTC 240
DB 181 CTGGCGGGGCCAACAACCTTGTGTATAGGCTGGGAGCCGTCGAAAACGAGGTCAACCTTC 240

DB 181 CTGGCGGGGCCAACAACCTTGTGTATAGGCTGGGAGCCGTCGAAAACGAGGTCAACCTTC 240
QY 241 ACACACCCCATTAACCAAAATTCATCATGAGTGCATGTGACGCCGACTGGAGGTGTACG 300
DB 241 ACACACCCCATTAACCAAAATTCATCATGAGTGCATGTGACGCCGACTGGAGGTGTACG 300
QY 301 AGCACCTGGGTGCTGGTGGGGGTCTTTCAGCTCTGGTGTGGGTATTTGCTTGAACA 360
DB 301 AGCACCTGGGTGCTGGTGGGGGTCTTTCAGCTCTGGTGTGGGTATTTGCTTGAACA 360
QY 361 GGCAGCGTGGTCAATGTGGTAGATCATCTTGTCCGGGGCGGAGCTATTGTTCCGAC 420
DB 361 GGCAGCGTGGTCAATGTGGTAGATCATCTTGTCCGGGGCGGAGCTATTGTTCCGAC 420
QY 421 AGGGAAGTCTCTACAGAGATTGATAGATGGAAGAGTGCAGCTTCCTCTTAC 480
DB 421 AGGGAAGTCTCTACAGAGATTGATAGATGGAAGAGTGCAGCTTCCTCTTAC 480
QY 481 ATCGAGCAGGGAATGACACTGCGGAGATTCAGCAAAAAGCGCTCGGGTTGCTGAG 540
DB 481 ATCGAGCAGGGAATGACACTGCGGAGATTCAGCAAAAAGCGCTCGGGTTGCTGAG 540
QY 541 ACAGCCACCAAGCAAGCGGAGCGCTGCTCCGTTGTGAGTCCAAAGTGGAGCCCTT 600
DB 541 ACAGCCACCAAGCAAGCGGAGCGCTGCTCCGTTGTGAGTCCAAAGTGGAGCCCTT 600
QY 601 GAGACCTTTCTGGGCGAAAACATGTGAACTTCATACAGCGGGAATACATTTAGCAGGC 660
DB 601 GAGACCTTTCTGGGCGAAAACATGTGAACTTCATACAGCGGGAATACATTTAGCAGGC 660
QY 661 TTGTCACTTGTGCTGGGAATCCCGCATTCATCATGATGAGTGGGTTTCAAGCTCTGTC 720
DB 661 TTGTCACTTGTGCTGGGAATCCCGCATTCATCATGATGAGTGGGTTTCAAGCTCTGTC 720
QY 721 ACTAGCCGCTCACCAACCAATTAACCTCCGTTTACATCTGAGGGGATGGGTAGCC 780
DB 721 ACTAGCCGCTCACCAACCAATTAACCTCCGTTTACATCTGAGGGGATGGGTAGCC 780
QY 781 GCCCAACTCGCTCCCGCAGTGTCTTCACTTTCATGAGCGCGGCAATTCCTGATGCG 840
DB 781 GCCCAACTCGCTCCCGCAGTGTCTTCACTTTCATGAGCGCGGCAATTCCTGATGCG 840
QY 841 GCTGTTGCGAGATGAGCTTGGGAAAGTGTCTTGTGACATCTTGGCGGGCTATGAGACA 900
DB 841 GCTGTTGCGAGATGAGCTTGGGAAAGTGTCTTGTGACATCTTGGCGGGCTATGAGACA 900
QY 901 GAGTGGAGAGCGGCTGTGGCTTTAAGGTCATGAGCGGGGAAATGCCCTCCACCGAG 960
DB 901 GAGTGGAGAGCGGCTGTGGCTTTAAGGTCATGAGCGGGGAAATGCCCTCCACCGAG 960
QY 961 GACTGTGTTAACTTAATCTCCCTGCACTCTCTCTGTGGTCCCTGGTGTGGGGTCTG 1020
DB 961 GACTGTGTTAACTTAATCTCCCTGCACTCTCTCTGTGGTCCCTGGTGTGGGGTCTG 1020
QY 1021 TGGCGAGGATATGCGTGGGACGTGGGTCCAGGGGAGGGGGCTGTGCAGTGATGAAC 1080
DB 1021 TGGCGAGGATATGCGTGGGACGTGGGTCCAGGGGAGGGGGCTGTGCAGTGATGAAC 1080
QY 1081 CGGCTGATAGGCTTCCGCTCGGGGTAAACAATTTTCCCGACGACATGATGTGCAGAG 1140
DB 1081 CGGCTGATAGGCTTCCGCTCGGGGTAAACAATTTTCCCGACGACATGATGTGTGCAGAG 1140
QY 1141 AGCGAGCGCGAGCAGTGTCACTAGATCTCTCCGACTTACTATGACCAACATGTTG 1200
DB 1141 AGCGAGCGCGAGCAGTGTCACTAGATCTCTCTCCGACTTACTATGACCAACATGTTG 1200
QY 1201 AAGAGCTTCAACCAATGATTAACGAGAGTGTCTCCAGCCCTGTCCGGCTGTGCTGA 1260
DB 1201 AAGAGCTTCAACCAATGATTAACGAGAGTGTCTCCAGCCCTGTCCGGCTGTGCTGA 1260
QY 1261 AGGAGTGTGAGATGATGATGACAGTGTGCTGACTTCAACAACCTGGCTCCAGTCC 1320
DB 1261 AGGAGTGTGAGATGATGATGACAGTGTGCTGACTTCAACAACCTGGCTCCAGTCC 1320

QY 1321 AAGCTCTGCGCGGATTAACGGAGATCCCTTTTCTCATGCGCAAGTGGGTCAAGGG 1380
 DB 1321 AAGCTCTGCGCGGATTAACGGAGATCCCTTTTCTCATGCGCAAGTGGGTCAAGGG 1380
 QY 1381 GTCTGGCGGGAGACGGCATCATGACACCACTGCTCATGTGGACACAGATCACCGA 1440
 DB 1381 GTCTGGCGGGAGACGGCATCATGACACCACTGCTCATGTGGACACAGATCACCGA 1440
 QY 1441 CATGTCAAAAACGGTTCATGAGGATCGTTGGGCTTAAGACCTGTAGTAACATGTGGCT 1500
 DB 1441 CATGTCAAAAACGGTTCATGAGGATCGTTGGGCTTAAGACCTGTAGTAACATGTGGCT 1500
 QY 1501 GGAACATTTCCCATCAACGATACACACGCGGCGCTGACGCGCTCCCGACGGCCAAAC 1560
 DB 1501 GGAACATTTCCCATCAACGATACACACGCGGCGCTGACGCGCTCCCGACGGCCAAAC 1560
 QY 1561 TATTCAGAGGCGCTGTGGCGGCTGCTGAGAGTACGTGAGGTTACGCGGATGGG 1620
 DB 1561 TATTCAGAGGCGCTGTGGCGGCTGCTGAGAGTACGTGAGGTTACGCGGATGGG 1620
 QY 1621 GATTTCCACTACGTGACGAGCATGACCACTGACAACTAAATGCCCGTCCAGGTTCCA 1680
 DB 1621 GATTTCCACTACGTGACGAGCATGACCACTGACAACTAAATGCCCGTCCAGGTTCCA 1680
 QY 1681 GCGCCCGAATTTCTTCAGAGAGTGTGGGGTGGCGCTGACAGGTAAGCTCCGAGGTGC 1740
 DB 1681 GCGCCCGAATTTCTTCAGAGAGTGTGGGGTGGCGCTGACAGGTAAGCTCCGAGGTGC 1740
 QY 1741 AAACCTCTCTACGAGGAGAGGTACATTCAGAGTGGGCTCAACCAATACCTGTTGGG 1800
 DB 1741 AAACCTCTCTACGAGGAGAGGTACATTCAGAGTGGGCTCAACCAATACCTGTTGGG 1800
 QY 1801 TGGCAGCTCCCATGCGAGCGCCCAACCGGATGTGACGTGCTCACTTCATGCTCACCGAC 1860
 DB 1801 TGGCAGCTCCCATGCGAGCGCCCAACCGGATGTGACGTGCTCACTTCATGCTCACCGAC 1860
 QY 1861 CCTCCCATCATCAAGCAGAGAGCGCTAAGCGGCTGGCGCGGAGGCTCTCCCGCTCC 1920
 DB 1861 CCTCCCATCATCAAGCAGAGAGCGCTAAGCGGCTGGCGCGGAGGCTCTCCCGCTCC 1920
 QY 1921 TTGGCAGCTCTTCAGCTAGCCAGTGTGTGCGCTTCTCTGAGGCGCATATCATTTACC 1980
 DB 1921 TTGGCAGCTCTTCAGCTAGCCAGTGTGTGCGCTTCTCTGAGGCGCATATCATTTACC 1980
 QY 1981 CAAATGATTTCCAGAGCGCTGACTCATGAGGCGCACTCTGTGGCGCATGAGATG 2040
 DB 1981 CAAATGATTTCCAGAGCGCTGACTCATGAGGCGCACTCTGTGGCGCATGAGATG 2040
 QY 2041 GGC 2043
 DB 2041 GGC 2043

RESULT 4
 LOCUS A28157 3750 bp DNA linear PART 04-AUG-1995
 DEFINITION PT-NANBH mRNA fragment from patent GB2239245.
 ACCESSION A28157 GI:1248640
 VERSION A28157.1 GI:1248640
 KEYWORDS
 SOURCE synthetic construct
 ORGANISM other sequences; artificial sequences.
 REFERENCE 1 (bases 1 to 3750)
 AUTHORS Post-transfusal non-A non-B hepatitis viral polypeptides
 JOURNAL Patent: GB 2239245-A 22-26-JUN-1991;
 FEATURES Location/Qualifiers
 1..3750
 /organism="synthetic construct"
 /mol_type="unassigned DNA"
 /db_xref="taxon:32630"

CDS
 /note="contig formed by cdna clones from 3' end"
 <1..3750
 /codon_start=1
 /transl_table=1
 /product="PT-NANBH polypeptide"
 /protein_id="CAA01932.1"
 /db_xref="GI:1248641"
 /translation="MEGVFTGLTHVDHFLSQTQKQDNFYLVAVQATVCARAQAP
 PSMQMMKTLRLKPTLNGPPLRLYLRLAGAVQNEVTLTHPILFKIEMACNADAEVLTST
 WLVGVLAALAAVLYLTGTSVIVTRIIISGRPAIVPREVLYQEDMEBCASLHPY
 IEQMQLEQFKQKGLQITKQAEAAAPVSKMRPLETFMKHMINPISGLQYL
 AGLSTLPGNPILASIAFTASVSPILTTQSTLLNLI LGGWVAQOLA PPSAASAPFAG
 IAGAIVSGIGLKVLDILAGYAGVAGLVLFKMGSGMPESTEDLVNLPILPFGA
 LVGVQACAILERHYGPGGAVQMMNRILAFPSRNHNSPTHYVESDAAAVTQILS
 DLITQILKRIHOWINDEQSTPSCGSMNRLCTYLADFKTMLOSILRLPGVP
 PFSQCRGKGVNRGGINQITSCCAQITGHAKNSMNI VEGKTSNMHGFPPINAY
 TTGCTPSPAPNYSALNRVAEEVEVTRVDFHYVISMITDNKPCQVPAPEFT
 EVDGVLHRYAPACKPLREBVTFOVGNQYLVGSQLPCEPDPDAVLITMLDPSHI
 TAEAKRLRARGSPPSLASASQASBSKATVITQNDFPDADILEANLIMRHMG
 DITRVSEKRVILDSFPLRAEBDEREVSPAEILRKSKEPMPAMPARDVNPPL
 LESKAPVYPPVHVCPLPPTKPTPIPEPRKRTVVLTESTVSSALAEIARFAGSS
 EBSAVDSGTATAPPPQSDDGAGSDVSYSSMPLEBGRPDLSDSMSIVSEBAG
 EDVCCSMSYMTGALITPCAEBSKLPINALSNLILRHNNVATITSRASQOKKV
 TFDRLQILDDHYQDVLKEMKAVSTVKAKLSEBAECKLTPHSAKSKRFGAKOVN
 LSKAIIHNRVSWEDLEBDETETPIDTTIMAKNEVFCVOPERGRRPARLIVPDDGVR
 VCEKMAVDVSTLPOAVMGSSYGFQYSGRVEFLVNAWMSKPTMCGAYDTRCEDS
 TYTENDIYVESITVCCDLAPBAROALISITRVLITGELTMSKONGCRASGV
 LTTSGNTITCTIKASACRAKLDCTMLVCGDGLVITCEBAGTQEDBASLRVTEA
 MTRVAPDPDPQPEYLDLELTSSSNVSVAHDAAGKRVYLYLRDP"

ORIGIN
 Query Match 100.0%; Score 2043; DB 6; Length 3750;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2043; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGAGAGGCGTCTTCAAGGCTCAACCACTGATGCCACTTCTGTCACCAACAAAG 60
 DB 1 TGGAGAGGCGTCTTCAAGGCTCAACCACTGATGCCACTTCTGTCACCAACAAAG 60
 QY 61 CAGGACGAGACAACTTCCCTTACTCTGTGGCGTACAGGCTACTGTGCGCTAGGGC 120
 DB 61 CAGGACGAGACAACTTCCCTTACTCTGTGGCGTACAGGCTACTGTGCGCTAGGGC 120
 QY 121 CAGGCCCACTCCATCATGAGATCAAAATGTGAAATGTCTCATTCGCGTAAAGCTACT 180
 DB 121 CAGGCCCACTCCATCATGAGATCAAAATGTGAAATGTCTCATTCGCGTAAAGCTACT 180
 QY 181 CTGCGGCGCCAAACCTTGTCTGTATAGCTGGAGCGCTCCAAACGAGGTCAACCTTC 240
 DB 181 CTGCGGCGCCAAACCTTGTCTGTATAGCTGGAGCGCTCCAAACGAGGTCAACCTTC 240
 QY 241 ACACACCCCATACCAAAATTCATATGAGCATGTCTACGCCGACTGAGGTCTGACG 300
 DB 241 ACACACCCCATACCAAAATTCATATGAGCATGTCTACGCCGACTGAGGTCTGACG 300
 QY 301 AGCAGCTGGTCTGTGGCGGGGCTCTTGAGCTGTGGCGGTATGCTTGAACAACA 360
 DB 301 AGCAGCTGGTCTGTGGCGGGGCTCTTGAGCTGTGGCGGTATGCTTGAACAACA 360
 QY 361 GGCAGCGTGTATGTGGGTAGATCATCTTGTCCGGCGGCGGCTATTGTTCCCGAC 420
 DB 361 GGCAGCGTGTATGTGGGTAGATCATCTTGTCCGGCGGCGGCTATTGTTCCCGAC 420
 QY 421 AAGGAAGTCTCTTACAGAGATTGATGATGAGAAAGTGGCGGTGCACTCCCTTAC 480
 DB 421 AAGGAAGTCTCTTACAGAGATTGATGATGAGAAAGTGGCGGTGCACTCCCTTAC 480
 QY 481 ATCGAGCGAGGAATGAGCTGCGAGCATCAAGCAAAAGCGCTGGGTGTCGACG 540
 DB 481 ATCGAGCGAGGAATGAGCTGCGAGCATCAAGCAAAAGCGCTGGGTGTCGACG 540
 QY 541 ACAGCCACCAAGCAGGAGCGCTGCTCCGCTGTGAGACTCAAGTGGCGAGCCCTT 600

Db 541 ACAGCCACCAAGCAAGGAGGCGCTGCTCCCGTGTGAGTCCAGTGGCAGCCCTT 600
 Qy 601 GAGACCTTCTGGGGCAACACATGTGAATTCATCAGCGGATACAGTACTTACAGGC 660
 Db 601 GAGACCTTCTGGGGCAACACATGTGAATTCATCAGCGGATACAGTACTTACAGGC 660
 Qy 661 TTGTCACTCTGCTGGGAAATCCCGGATTTGATCACTGATGGCTTACAGCCTCTGTC 720
 Db 661 TTGTCACTCTGCTGGGAAATCCCGGATTTGATCACTGATGGCTTACAGCCTCTGTC 720
 Qy 721 ACTACCCGCTACCAACCAATCTACCTCCGCTTAAATCATCTGCGGGGATGGGTAAGC 780
 Db 721 ACTACCCGCTACCAACCAATCTACCTCCGCTTAAATCATCTGCGGGGATGGGTAAGC 780
 Qy 781 GCCCAACTGCTCCCCCAGTGTCTTCACTTCTGATGAGCGCGGCAATTTCTGTGTC 840
 Db 781 GCCCAACTGCTCCCCCAGTGTCTTCACTTCTGATGAGCGCGGCAATTTCTGTGTC 840
 Qy 841 GCTGTGGCAGATAGGCTTGGGAAAGTGTGTGAGATCTTGGCGGGCTATGAGAGA 900
 Db 841 GCTGTGGCAGATAGGCTTGGGAAAGTGTGTGAGATCTTGGCGGGCTATGAGAGA 900
 Qy 901 GAGAGGCAAGGCGCGCTGTGGCTTTAAGTCAAGAGCGGAAATGCCCTCCACCGAG 960
 Db 901 GAGAGGCAAGGCGCGCTGTGGCTTTAAGTCAAGAGCGGAAATGCCCTCCACCGAG 960
 Qy 961 GACCTGTAACTTCTCCCTGCAATCTCTCTCTGTGTGCTCGTGGTCTGCGGGTCTGT 1020
 Db 961 GACCTGTAACTTCTCCCTGCAATCTCTCTCTGTGTGCTCGTGGTCTGCGGGTCTGT 1020
 Qy 1021 TGGCAGCGATATCTGCTGCGACCTGTGGTCCAGGGGAGGGGCTGTGCAATGATGAC 1080
 Db 1021 TGGCAGCGATATCTGCTGCGACCTGTGGTCCAGGGGAGGGGCTGTGCAATGATGAC 1080
 Qy 1081 CGGCTGATAGCGTGTGGCTCGGGGGTAAACAATTTCCCAACGACTATGTGCAAGAG 1140
 Db 1081 CGGCTGATAGCGTGTGGCTCGGGGGTAAACAATTTCCCAACGACTATGTGCAAGAG 1140
 Qy 1141 AGCGAGCGCGACGATGTCACTAGATCTCTCCGACTTACTATCAACCACTGTGTG 1200
 Db 1141 AGCGAGCGCGACGATGTCACTAGATCTCTCCGACTTACTATCAACCACTGTGTG 1200
 Qy 1201 AAGAGGCTTCCACGATGATTAACAAGAGCTGCTCCAGCCCTGCTCGGGCTGTGCTTA 1260
 Db 1201 AAGAGGCTTCCACGATGATTAACAAGAGCTGCTCCAGCCCTGCTCGGGCTGTGCTTA 1260
 Qy 1261 AAGGAGTGTGGGATCTGATTAAGCAAGTGTGGCTGATTAAGAAGCTGCTCAAGTCC 1320
 Db 1261 AAGGAGTGTGGGATCTGATTAAGCAAGTGTGGCTGATTAAGAAGCTGCTCAAGTCC 1320
 Qy 1321 AAGCTCTGCGCGATTAACCGGAGTCCCTTTTCTCATGTGCCAACTGTGGTAAAGGG 1380
 Db 1321 AAGCTCTGCGCGATTAACCGGAGTCCCTTTTCTCATGTGCCAACTGTGGTAAAGGG 1380
 Qy 1381 GTCTGGCGGGAGAGCGGATTCATGACACCTGTCTCATGTGAGCAAGATCACCGGA 1440
 Db 1381 GTCTGGCGGGAGAGCGGATTCATGACACCTGTCTCATGTGAGCAAGATCACCGGA 1440
 Qy 1441 CATGTCAAAAAGCTTCCATGAGATCTGTGGGCTTAAGACCTGTATGATAAGTGGGAT 1500
 Db 1441 CATGTCAAAAAGCTTCCATGAGATCTGTGGGCTTAAGACCTGTATGATAAGTGGGAT 1500
 Qy 1501 GGAACATTTCCCATCAAGCATACACAGGAGCCCTGTGACGCCCTCCCAAGCCCAAC 1560
 Db 1501 GGAACATTTCCCATCAAGCATACACAGGAGCCCTGTGACGCCCTCCCAAGCCCAAC 1560
 Qy 1561 TAATTCAGGGGCTGTGTGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1620
 Db 1561 TAATTCAGGGGCTGTGTGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1620
 Qy 1621 GATTTTCATGATGAGCATGACCACTGACAAAGTAAATGCCCCGTGCGAGTTTCCA 1680
 Db 1621 GATTTTCATGATGAGCATGACCACTGACAAAGTAAATGCCCCGTGCGAGTTTCCA 1680

Qy 1681 GCCCCGGAATTCCTOACAGAAATGATGGGTGGGCTGTGACAGGTACGCTTCGGCGTTC 1740
 Db 1681 GCCCCGGAATTCCTOACAGAAATGATGGGTGGGCTGTGACAGGTACGCTTCGGCGTTC 1740
 Qy 1741 AAACCTCTTACGGGAGGAGGTGACATTCAGAGTGGGGCTCAACCAATACCTGGTTGG 1800
 Db 1741 AAACCTCTTACGGGAGGAGGTGACATTCAGAGTGGGGCTCAACCAATACCTGGTTGG 1800
 Qy 1801 TCGAGCTCCCATCGAGCCGAAACCGGATGAGTGTCACTTCCATGCTCACCGAC 1860
 Db 1801 TCGAGCTCCCATCGAGCCGAAACCGGATGAGTGTCACTTCCATGCTCACCGAC 1860
 Qy 1861 CCTTCCCATOACAGAGAGAGCGCTTAAGGCGAGGCTGGCCAGGGGGTCTTCCCCCTTC 1920
 Db 1861 CCTTCCCATOACAGAGAGAGCGCTTAAGGCGAGGCTGGCCAGGGGGTCTTCCCCCTTC 1920
 Qy 1921 TTGGCAGCTCTTACGCTAGCGAGTGTCTGCGCTTCTCGAAGGCGACATATTAAC 1980
 Db 1921 TTGGCAGCTCTTACGCTAGCGAGTGTCTGCGCTTCTCGAAGGCGACATATTAAC 1980
 Qy 1981 CAAATGACTTCCAGAGCGTGAACCTCATTCAGAGCCACTCTGTGGCGGATGAGATG 2040
 Db 1981 CAAATGACTTCCAGAGCGTGAACCTCATTCAGAGCCACTCTGTGGCGGATGAGATG 2040
 Qy 2041 GGC 2043
 Db 2041 GGC 2043
 RESULT 5
 ARI144051
 LOCUS ARI144051 3750 bp DNA linear PAT 08-AUG-2001
 DEFINITION Sequence 22 from patent US 6210675.
 ACCESSION ARI144051
 VERSION ARI144051.1 GI:15105918
 KEYWORDS
 SOURCE
 ORGANISM
 UNKNOWN.
 Unclassified.
 REFERENCE 1 (bases 1 to 3750)
 AUTHORS Highfield, P.Edmund., Rodgers, B.Colin., Tedder, R.Seton. and
 TITLE Barbara, J. Anthony, James.
 JOURNAL PT-NAMH hepatitis polypeptides
 PATENT: US 6210675-A 22 03-Apr-2001;
 FEATURES
 source 1..3750
 /organism="unknown"
 /mol_type="unassigned DNA"
 ORIGIN
 Query Match 100.0%; Score 2043; DB 6; Length 3750;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2043; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 TGGAGGGCGTCTTCAAGGCTCAACCACTGATGTCCTTCTGTCCAAACAAAG 60
 Db 1 TGGAGGGCGTCTTCAAGGCTCAACCACTGATGTCCTTCTGTCCAAACAAAG 60
 Qy 61 CAGGAGAGAGCAATTTCCCTCACTGGTGGGTACAGGCTACGTGTGGCTAGGGCC 120
 Db 61 CAGGAGAGAGCAATTTCCCTCACTGGTGGGTACAGGCTACGTGTGGCTAGGGCC 120
 Qy 121 CAGGCCCATCTCATCTATGAGTGAAGTGTCTCATACGCTAAAGCTTACT 180
 Db 121 CAGGCCCATCTCATCTATGAGTGAAGTGTCTCATACGCTAAAGCTTACT 180
 Qy 181 CTGCGCGGCGCAACACCTTGTGTATAGGCTGGAGCGTTCACAAAGAGGTACCTTC 240
 Db 181 CTGCGCGGCGCAACACCTTGTGTATAGGCTGGAGCGTTCACAAAGAGGTACCTTC 240
 Qy 241 ACACACCCATTAACCAATTCATCATGTGATGATGTCAAGCGACCTGAAGTCTCTACG 300
 Db 241 ACACACCCATTAACCAATTCATCATGTGATGATGTCAAGCGACCTGAAGTCTCTACG 300

| | | | |
|----|------|--|------|
| Db | 241 | ACACACCCCATPACCAAAATTCATCATGGCATGCAATGTCAGGCCGACCTG9AAGGTGCTCACG | 300 |
| Qy | 301 | AGCACTTGGGTGTCTGTGGGCGGGGTCTTGTGACGCTCTGTGCTGTATTTGCTTGACAACA | 360 |
| Db | 301 | AGCACTTGGGTGTCTGTGGGCGGGGTCTTGTGACGCTCTGTGCTGTATTTGCTTGACAACA | 360 |
| Qy | 361 | GGCAGCGTGTGCTATTGTGGGTAAAGATCAATCTTGTGCGGGGCGGCGGTATTTGTTCCGAC | 420 |
| Db | 361 | GGCAGCGTGTGCTATTGTGGGTAAAGATCAATCTTGTGCGGGGCGGCGGTATTTGTTCCGAC | 420 |
| Qy | 421 | AGGGAAGTCTCTACCAAGAGTTTCAGATAGAAAGATGCGCCTCGCACCTCCCTTAC | 480 |
| Db | 421 | AGGGAAGTCTCTACCAAGAGTTTCAGATAGAAAGATGCGCCTCGCACCTCCCTTAC | 480 |
| Qy | 481 | ATCGAGCAGGAAATGACGCTTCGCGAGCAGTTCAAGCAAAAAGCGCTCGGTTGTGTCAG | 540 |
| Db | 481 | ATCGAGCAGGAAATGACGCTTCGCGAGCAGTTCAAGCAAAAAGCGCTCGGTTGTGTCAG | 540 |
| Qy | 541 | ACACCCACCAAGCAAGCGGAGGCGGTCTCCCGTGTGGAGTCCAAATGGCGAGCCCTT | 600 |
| Db | 541 | ACACCCACCAAGCAAGCGGAGGCGGTCTCCCGTGTGGAGTCCAAATGGCGAGCCCTT | 600 |
| Qy | 601 | GAGACTTTCTGGGCGAAAACAATGTGAAATTTCAATCAGCGGATPACATTAATTAGCAGGC | 660 |
| Db | 601 | GAGACTTTCTGGGCGAAAACAATGTGAAATTTCAATCAGCGGATPACATTAATTAGCAGGC | 660 |
| Qy | 661 | TTGTTCACACTGTGCTGGGAAATCCCGCATGTCATCTGATGTGGCTTGACAGCCTGTGC | 720 |
| Db | 661 | TTGTTCACACTGTGCTGGGAAATCCCGCATGTCATCTGATGTGGCTTGACAGCCTGTGC | 720 |
| Qy | 721 | ACTAGCCCGCTTCACCAACCAATCTAACCTCTGTGCTTAAATCTGTGGGGGATGGGTACC | 780 |
| Db | 721 | ACTAGCCCGCTTCACCAACCAATCTAACCTCTGTGCTTAAATCTGTGGGGGATGGGTACC | 780 |
| Qy | 781 | GCCCAACTCGCTCCCGCCAGTGTGCTTCAAGCTTTGTAAGGCGCGGCAATGTGTGTGCG | 840 |
| Db | 781 | GCCCAACTCGCTCCCGCCAGTGTGCTTCAAGCTTTGTAAGGCGCGGCAATGTGTGTGCG | 840 |
| Qy | 841 | GCTGTGGCAGACATAGGCTTGGGAAGGTGCTTGTGACATCTTGTGGCGGCTATGAGACA | 900 |
| Db | 841 | GCTGTGGCAGACATAGGCTTGGGAAGGTGCTTGTGACATCTTGTGGCGGCTATGAGACA | 900 |
| Qy | 901 | GGAATGCGAGGCGGCTCTGTGGCCTTTAAAGTCAATGACGGCGGAATGCCCTCCACCGAG | 960 |
| Db | 901 | GGAATGCGAGGCGGCTCTGTGGCCTTTAAAGTCAATGACGGCGGAATGCCCTCCACCGAG | 960 |
| Qy | 961 | GACCTGGTTAACTTAATCTCCCTGCGCACTCTCTCTCTGCTGTGCGCTGTGGGTGCTGTG | 1020 |
| Db | 961 | GACCTGGTTAACTTAATCTCCCTGCGCACTCTCTCTCTGCTGTGCGCTGTGGGTGCTGTG | 1020 |
| Qy | 1021 | TGCGCAGCGATATGCGCTCGGCAAGTGGGTCAAGGGAGGGGGGCTGTGTCAGTGAATGAC | 1080 |
| Db | 1021 | TGCGCAGCGATATGCGCTCGGCAAGTGGGTCAAGGGAGGGGGGCTGTGTCAGTGAATGAC | 1080 |
| Qy | 1081 | CGGCTGATAGCGTTGCTCGCTCGCGGGGTAAACATGTTTCCCCACGCACTAATGTGCGAAG | 1140 |
| Db | 1081 | CGGCTGATAGCGTTGCTCGCTCGCGGGGTAAACATGTTTCCCCACGCACTAATGTGCGAAG | 1140 |
| Qy | 1141 | AGGCAAGCGGAGGACGTTGTCACATCAAGTCTCTCCGACTTATCAATCAACCAACTGTGG | 1200 |
| Db | 1141 | AGGCAAGCGGAGGACGTTGTCACATCAAGTCTCTCCGACTTATCAATCAACCAACTGTGG | 1200 |
| Qy | 1201 | AAGAGGCTCCACAGTGAATTAAGAGAGCTGTCCAGCGCCTGTGCTCGGCTCGTGTGCTA | 1260 |
| Db | 1201 | AAGAGGCTCCACAGTGAATTAAGAGAGCTGTCCAGCGCCTGTGCTCGGCTCGTGTGCTA | 1260 |
| Qy | 1261 | AGGATGTTTGGGACTGGAATGCAACAGTTTTGTGCTGACTTCAAGACTGTGCTCAAGTCC | 1320 |
| Db | 1261 | AGGATGTTTGGGACTGGAATGCAACAGTTTTGTGCTGACTTCAAGACTGTGCTCAAGTCC | 1320 |
| Qy | 1321 | AAAGTCTCCGCGCGCATTAACCGGAGTCCCTCTTTTTCATATGCAACATGTGGTAAAGGGG | 1380 |
| Db | 1321 | AAAGTCTCCGCGCGCATTAACCGGAGTCCCTCTTTTTCATATGCAACATGTGGTAAAGGGG | 1380 |

| | | | |
|------------|---|--|-----------------|
| Qy | 1381 | GTCTGCGGGGAGACGGCATCATGACAACCACTGCTCATGTGAGACAGATACCGGA | 1440 |
| Db | 1381 | GTCTGCGGGGAGACGGCATCATGACAACCACTGCTCATGTGAGACAGATACCGGA | 1440 |
| Qy | 1441 | CATGTCAAAAACGGTTCATGAGGATCGTTGGGCTTAAGACCTGTAGTAACATGTGGCAT | 15000 |
| Db | 1441 | CATGTCAAAAACGGTTCATGAGGATCGTTGGGCTTAAGACCTGTAGTAACATGTGGCAT | 15000 |
| Qy | 1501 | GGAACTATCCCAATCAACGATTAACAACGCGGCCCCCTGCAACGCCCTCCCAAGGCCCAAAAC | 15666 |
| Db | 1501 | GGAACTATCCCAATCAACGATTAACAACGCGGCCCCCTGCAACGCCCTCCCAAGGCCCAAAAC | 15666 |
| Qy | 1561 | TATTCACAAGGCGCTGTGGCGGGTGGCTGCTGAGAGATACGTGAGATTACCGCGGGTGGGG | 16220 |
| Db | 1561 | TATTCACAAGGCGCTGTGGCGGGTGGCTGCTGAGAGATACGTGAGATTACCGCGGGTGGGG | 16220 |
| Qy | 1621 | GATTTCCACTACGTGACGAGCATGACCACTGACACAGTAATAATGCCCCGTCCAGTTTCCA | 16880 |
| Db | 1621 | GATTTCCACTACGTGACGAGCATGACCACTGACACAGTAATAATGCCCCGTCCAGTTTCCA | 16880 |
| Qy | 1681 | GCCCCCGAATTCCTTACAGAAAGTGAGATGGGGTGGCTGCAACAGTACGCTCCGGCGGTGC | 17440 |
| Db | 1681 | GCCCCCGAATTCCTTACAGAAAGTGAGATGGGGTGGCTGCAACAGTACGCTCCGGCGGTGC | 17440 |
| Qy | 1741 | AAACCTCTCCCTACGCGGAGAGAGTCACTTCAGAGTCCGGGCTCAACCAATACCTTGTTGGG | 18000 |
| Db | 1741 | AAACCTCTCCCTACGCGGAGAGAGTCACTTCAGAGTCCGGGCTCAACCAATACCTTGTTGGG | 18000 |
| Qy | 1801 | TCGCAGCTCCCATGCGAGCCCGAAACCGGATGTAGCAAGTCTCACTTCATGCTCACCGAC | 18660 |
| Db | 1801 | TCGCAGCTCCCATGCGAGCCCGAAACCGGATGTAGCAAGTCTCACTTCATGCTCACCGAC | 18660 |
| Qy | 1861 | CCCTCCCAACATCAAGAGAGAGACGGCTTAACGCGAGGTGGCCACGGGGGTCTCCCCCTCC | 19220 |
| Db | 1861 | CCCTCCCAACATCAAGAGAGAGACGGCTTAACGCGAGGTGGCCACGGGGGTCTCCCCCTCC | 19220 |
| Qy | 1921 | TTGGCCAGACTCTTTCAGCTAGACCAAGTTGTTCTGCGCTTCTCTGAAGGCGACATACATTACC | 19880 |
| Db | 1921 | TTGGCCAGACTCTTTCAGCTAGACCAAGTTGTTCTGCGCTTCTCTGAAGGCGACATACATTACC | 19880 |
| Qy | 1981 | CAAAATGACTTCCAGACGCTGACCTCATGAGGCCAACTCTCTGTGGCGGCATGAGATG | 20440 |
| Db | 1981 | CAAAATGACTTCCAGACGCTGACCTCATGAGGCCAACTCTCTGTGGCGGCATGAGATG | 20440 |
| Qy | 2041 | GGC 2043 | |
| Db | 2041 | GGC 2043 | |
| RESULT 6 | | | |
| LOCUS | A76577 | 7065 bp | DNA |
| DEFINITION | Sequence 9 from Patent W09317110. | linear | PAT 19-OCT-1999 |
| ACCESSION | A76577 | | |
| VERSION | A76577.1 | | |
| KEYWORDS | GI:6088488 | | |
| SOURCE | unidentified | | |
| ORGANISM | unidentified | | |
| REFERENCE | 1 (bases 1 to 7065) | | |
| AUTHORS | Rodgers,B.C. and Parker,D. | | |
| TITLE | A RECOMBINANT HEPATITIS C VIRUS POLYPEPTIDE | | |
| JOURNAL | Patent: WO 931710-A 9 02-SEP-1993; | | |
| FEATURES | WELLCOME FOUND (GB) | | |
| source | Location/Qualifiers | | |
| | 1..7065 | | |
| | /organism="unidentified" | | |
| | /mol_type="unassigned DNA" | | |
| | /db_xref="taxon:32644" | | |
| | <1..7065 | | |
| | /note="Encoded putative PT-NANBH non-structural proteins" | | |
| | /codon_start=1 | | |

/product="P1-NANBH coding sequence"
/protein_id="CAB58625.1"
/db_xref="GI:6088489"
translation="SCGAGVFIQLMPLTSPYKVFIAKLIMLQYPIITRAEHLQVW
VPLNVRGGADAILLACVHBEPLTSGIKILLGPMLOAGITRVPYRQGL
IRACLVKRVGGHYVOMALMKLALITGVYVDHLTPLRDMHAKRLDAVAEVEP
SDMERVITWGTDTAACGDIILGLPYVARGBEITFLGRPDSIEGGKMLAPITAYSQ
QTRGLGCTITSITGRDNQVEGEVQVSTATOSFLATVCSVCVCTHAGVITVYSG
PKGPVIMYTVNDODLVGMAPPGARSLTPCGSGDYLVRHADVLPVRGDSRG
SLSPRPISLYLKSGSGPLCPGSHAVGIFRAVCTRGAKVADEIIPESMETVYRSP
VETDSSPPAVPOSFOVALHAPVSGSKSTRPAAYAAQGYVLVNPVALAFGA
YMSKAGVDPNTRSGVTRITTTGAPIITVYSGKFLADGSGGAYIIICDBEHSIDST
SILIGCTVLDQAEHTRGARIYVATATPESVTVPHNPBEVALNMTGEPIPRGKALPI
ETIKGRHLIFCHSKKCDLAKLVGGINAVAYRGADVAPASGQVVAVATDAL
MIFGTGDPDSVIDCNTVQIVDFSLDPEFIETITVPDASRSQRGRTGRARDI
YRFVAPGRSPGMDPSVLCCEYDAGCAAMEYLPATSESTRALAYPNTDGLPCQDHL
FVESVETGLTHDAHFLSOTKQAGNPFVLVAYOATVCRABAPPSDQMKCIRL
KPTLRGPDLIYRLGAVONEVTLTHPIKFTMACMSADLEVTSTWVYGVATLAA
YCLTTGSVIVYGRITLSGRPAIVPDRBVIYQFDEMEBCASHPIITEGMLAEFPKA
KALGLLOTATKQAEAAAPVESKMRALETFMKHMNFTSGIQLYLAGSLTEGNTPA
SLMAPTASTPLTOSTLLNII LGGWAAQLAPPSAASFAGIAGAVAGSILGR
VLVDILAGYAGVALVAFKVSGEMPESTEDLVNLLPILSPGALVGVCAILIR
HYVQEGAVVOMNRIIAPASRGNHVSPTHYVPESDAARVQIISDLITTOILKHLH
WNEDESPVQSMNLRDWMDICTYLABPKTMLQSKLPRLBGVPRSCQDRGYKVMY
GDGIMQTTSCGAQITGHVKGSMRIYGEKTSNMHGFPIINATITTEGCTPSRAPHY
SRALMRVAEEVENVTRVGDFFHYVTSMITDNVCPQVPAPPEFTEVGVKRLHRYPA
CPULREBVTQVGLNOYLVGSQLPCEBPDVAVLTMLTDPSHITABKAKRLIARPS
PPLSASSASOLASPSKATYITONDFPADLIEMNLMRHMGDITRVESENVLVI
LDSFPLRABEBEREVSVAEILIRKSKRPMPAMPARDPNPILLESKMDYVDPY
VHGCLPPTKPTPIPPRKRPTVULTESSVALAELATKAGASESISGVSTAP
PDQSDDGAGSDVESISMPLEBEGPDPDSDGSWSVSEBAEDVYCSMSTTTP
GALITPCAEBESKLPINALSNSLRRHNNVAVTISRASQKRYKTPRLQILDHYQ
DYLKMKAKASTYKAKLLSVEBAKLTPEHSAKSRFGAKGNVRLSALNHSIVM
EDLEDTETPIDTITIMAKNEVFCVQPERGRKPARLIYFPDGVYVCEMALYDVTST
LPQAVMGSSYGFQYSGPQGVFEFLVNAKMSKTPMGAFAVTRFSDVTEINDIREDVET
AKLLSVEBAKLTPEHSAKSRFGAKGNVRLSALNHSIVMEDLEDTETPIDTITIMAK
IMAKNEVFCVQPERGRKPARLIYFPDGVYVCEMALYDVTSTPOAVMGSSYGFQY
SPQGVFEFLVNAKMSKTPMGAFAVTRFSDVTEINDIREDVETAKLLSVEBAKLTPEH
RSLETRLYIGPLTNSKQNGCGRCRASGVLTSCGNLTCTYLASACRAKALQDC
TMLVCGDILVVICESAGTOEDASLVEFEMTRRSPAGDPQPEYDELELTSCSN
VSVADHAGSKRYVLYTRDPTPLASARAMEATARTVMSGLNIIYVAPITVYRMLMT
HPSSTILAOEOLKLDQOIVGACSTIEPDIPOIIOIRHGISAKHSISGSEINRV
ASCRLKGVPLLRVNRHRAVRAKILSGBRAAICGRILPMVANKTKIKLPTITAAAS
RDLISGMFVAGSGDIIYHSVSRARPRMFMCLLLLSVGVLYLLPNR"
7063..7065
/function="Stop codon"

ORIGIN misc_signal
Query Match 99.6%; Score 2035; DB 6; Length 7065;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2038; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 TGGGAGGGGCTCTCA CAGGCGTCAACCGAGTGGAGTGGCCACTCTCTGTCACCAACAAG 60
DB 2224 TGGGAGGGGCTCTCA CAGGCGCTCAACCGAGTGGAGTGGCCACTCTCTGTCACCAACAAG 2283
QY 61 CAGGAGAGAGAACTTCCCTTA CTGTGCGGTACCAAGCTTACTGTGTGCGCTAGGCGC 120
DB 2284 CAGGAGAGAGAACTTCCCTTA CTGTGCGGTACCAAGCTTACTGTGTGCGCTAGGCGC 2343
QY 121 CAGGCGCCCACTTCATCGAGATCAATGTGGAAGTGTCTCATACGCGTTAAAGCTTACT 180
DB 2344 CAGGCGCCCACTTCATCGAGATCAATGTGGAAGTGTCTCATACGCGTTAAAGCTTACT 2403
QY 181 CTGGCGGGGCGCAACCCCTTGTGTATAGCTGGGAGCGGTCCAAACGAGGTCAACCCCTC 240
DB 2404 CTGGCGGGGCGCAACCCCTTGTGTATAGCTGGGAGCGGTCCAAACGAGGTCAACCCCTC 2463
QY 241 ACACACCCCATTAACAAATTCATCATGCGATGCGATGTCACCGACCTGAGAGGTGCTCAG 300
DB 2464 ACACACCCCATTAACAAATTCATCATGCGATGCGATGTCACCGACCTGAGAGGTGCTCAG 2523
QY 301 AGCACCCTGGGTGCTGGTGGCGGGGTCCTTTCAGCTTCGCGTGCATTTGCTTGAACA 360

DB 2524 AGCACCCTGGGTGCTGGTGGCGGGGTCCTTTCAGCTTCGCGTGCATTTGCTTGAACA 2583
QY 361 GGCACGCTGTGATCTGTGGGTAGATCATCTTGTCCGGGCGGCGCTATTGTTCCGAC 420
DB 2584 GGCACGCTGTGATCTGTGGGTAGATCATCTTGTCCGGGCGGCGCTATTGTTCCGAC 2643
QY 421 AGGGAATCTCTTACACAGAGTTCATAGATGGAAGAGTGCCTGCCACTCCCTTAC 480
DB 2644 AGGGAATCTCTTACACAGAGTTCATAGATGGAAGAGTGCCTGCCACTCCCTTAC 2703
QY 481 ATCGACGAGGGAATCAGCTCGCCGAGCATTTCAAGAAAAGCGCTGGGTGCTGCAG 540
DB 2704 ATCGACGAGGGAATCAGCTCGCCGAGCATTTCAAGAAAAGCGCTGGGTGCTGCAG 2763
QY 541 ACAGCCACCAAGCAAGCGAGGCGCTCTCCCGTGTGAAGTCCAACTGGCGAGCCCTT 600
DB 2764 ACAGCCACCAAGCAAGCGAGGCGCTCTCCCGTGTGAAGTCCAACTGGCGAGCCCTT 2823
QY 601 GAGACTTCTGGGCGAAACACATGTGAACTTCATCAGCGGATACATGACTTACAGGC 660
DB 2824 GAGACTTCTGGGCGAAACACATGTGAACTTCATCAGCGGATACATGACTTACAGGC 2883
QY 661 TTGTCCACTGCTGGGGAATCCGCGCATGTGCATCATGAGGCTTCAAGCCCTCTGC 720
DB 2884 TTGTCCACTGCTGGGGAATCCGCGCATGTGCATCATGAGGCGTTCACAGCTCTGTC 2943
QY 721 ACTAGCCGCTCACCACCAACTTACCTCTCTGCTTAACATCTTGGGGGATGGGTAGCC 780
DB 2944 ACTAGCCGCTCACCACCAACTTACCTCTCTGCTTAACATCTTGGGGGATGGGTAGCC 3003
QY 781 GCCCAACTGCTCCCCAGATGCTCTTCACTTGAAGGCGCGGCACTTGTCTGTGCG 840
DB 3004 GCCCAACTGCTCCCCAGATGCTCTTCACTTGAAGGCGCGGCACTTGTCTGTGCG 3063
QY 841 GCTGTGACAGATAGGCTTGGGAAAGGTGCTTGGAATCTTGGCGGCTATGAGACA 900
DB 3064 GCTGTGACAGATAGGCTTGGGAAAGGTGCTTGGAATCTTGGCGGCTATGAGACA 3123
QY 901 GGAAGTGGAGGCGGCGCTGCTGAGCTTTAAGTTCATGAGCGGGAATGCCCTTCA 960
DB 3124 GGAAGTGGAGGCGGCGCTGCTGAGCTTTAAGTTCATGAGCGGGAATGCCCTTCA 3183
QY 961 GACTGTGTTAACTTACTCCCTGCATCTCTCTCTGCTGCTGTGTGCGGGTCTGTG 1020
DB 3184 GACTGTGTTAACTTACTCCCTGCATCTCTCTCTGCTGCTGTGTGCGGGTCTGTG 3243
QY 1021 TGGCAGGAGATATGCGTCCGAGCGTGGGTCAGGGGAGGGGCTGTGAGTGAAG 1080
DB 3244 TGGCAGGAGATATGCGTCCGAGCGTGGGTCAGGGGAGGGGCTGTGAGTGAAG 3303
QY 3304 CGGCTGATAGGCTTGCTGCTCGGGGTTAACATATGTTCCCAAGCAGCATATGTGCA 3363
QY 1141 AGCGACCGCGAGACGTGTCTCATGATTCCTCTCGACCTTATCATACCACTTGTG 1200
DB 3364 AGCGACCGCGAGACGTGTCTCATGATTCCTCTCGACCTTATCATACCACTTGTG 3423
QY 1201 AAGAGGCTCCACCAATGATTTAAAGAGACGTGCTCAAGCCCTGTCCGGGCTCGGGCTA 1260
DB 3424 AAGAGGCTCCACCAATGATTTAAAGAGACGTGCTCAAGCCCTGTCCGGGCTCGGGCTA 3483
QY 1261 AGGAGTGTGAGACTGATATGCAACAGTTTGTGCTGACTTCAAGACCTGCTCCAGTCC 1320
DB 3484 AGGAGTGTGAGACTGATATGCAACAGTTTGTGCTGACTTCAAGACCTGCTCCAGTCC 3543
QY 1321 AAGCTCTGCGCGATTAACGGAGAGTCCCTTTTCTCATGCAACGTGAGTAAAGGGG 1380
DB 3544 AAGCTCTGCGCGATTAACGGAGAGTCCCTTTTCTCATGCAACGTGAGTAAAGGGG 3603
QY 1381 GTCTGGCGGGAGAGCGGATCATGCAACACCTGCTCATGTGAGGACAGATCACCGGA 1440


```

Db      3604 GTCTGGCGGGAGACGGCATCATGACACCACTGCTCATGTGAGACACAGATCACCGGA 3663
Qy      1441 CATGTCAAAACGCTTCATGAGATCGTTGGGCGCTAAGACCTGTAGTAACATGTGGCAT 1500
Db      3664 CATGTCAAAACGCTTCATGAGATCGTTGGGCGCTAAGACCTGTAGTAACATGTGGCAT 3723
Qy      1501 GGAACATTTCCCATCAACGATACACACGAGGCGCCCTGACGCGCTCCCGACGCGCAAC 1560
Db      3724 GGAACATTTCCCATCAACGATACACACGAGGCGCCCTGACGCGCTCCCGACGCGCAAC 3783
Qy      1561 TATTTCCAGGCGCTGTGGCGGGTGGTGTGTGAGAGATGATGTAAGTTACGCGGGTGGG 1620
Db      3784 TATTTCCAGGCGCTGTGGCGGGTGGTGTGTGAGAGATGATGTAAGTTACGCGGGTGGG 3843
Qy      1621 GATTTCCATGAGTACGAGCATGACCACTGCAACAGTAAATGCGCGCGGACAGGTTCCA 1680
Db      3844 GATTTCCATGAGTACGAGCATGACCACTGCAACAGTAAATGCGCGCGGACAGGTTCCA 3903
Qy      1681 GCGCCGGAATCTTCAACAGAAATGATGAGGCTGCGGCTGACAGGTAACGCTCCGCGTGC 1740
Db      3904 GCGCCGGAATCTTCAACAGAAATGATGAGGCTGCGGCTGACAGGTAACGCTCCGCGTGC 3963
Qy      1741 AAACCTCTCTACGGGAGAGAGTCACTTCCAGTGGGCTCAACCAATACCTGTTGGG 1800
Db      3964 AAACCTCTCTACGGGAGAGAGTCACTTCCAGTGGGCTCAACCAATACCTGTTGGG 4023
Qy      1801 TCGAGCTCCCATGAGGCGCGAAGCGATGAGCATGCTGCTCAATGCTCAACCGAC 1860
Db      4024 TCGAGCTCCCATGAGGCGCGAAGCGATGAGCATGCTGCTCAATGCTCAACCGAC 4083
Qy      1861 CCTCCCATCATCAACGAGAGAGCGCTAAGCGAGCTGCGAGGGGCTCTCCCGCTCC 1920
Db      4084 CCTCCCATCATCAACGAGAGAGCGCTAAGCGAGCTGCGAGGGGCTCTCCCGCTCC 4143
Qy      1921 TTGGCAGCTCTTCAAGCTAGCAAGTGTCTGCGCTTCTCTCAAGGCGCATATACCT 1980
Db      4144 TTGGCAGCTCTTCAAGCTAGCAAGTGTCTGCGCTTCTCTCAAGGCGCATATACCT 4203
Qy      1981 CAAATGACTTCCGAGAGCGCTACCTCATGAGGCGCACTCTGAGGCGGATAGATG 2040
Db      4204 CAAATGACTTCCGAGAGCGCTACCTCATGAGGCGCACTCTGAGGCGGATAGATG 4263
Qy      2041 GGC 2043
Db      4264 GGC 4266

```

```

RESULT 7
A32203 3750 bp DNA linear PART 10-DEC-1996
LOCUS NANBH PT polyprotein fragment.
DEFINITION A32203
ACCESSION A32203
VERSION A32203.1 GI:1926564
KEYWORDS
SOURCE Non-A, non-B hepatitis virus
ORGANISM Non-A, non-B hepatitis virus
REFERENCE Viruses; unclassified viruses.
AUTHORS 1 (bases 1 to 3750)
JOURNAL Patent: FR 2655990-A 22 21-JUN-1991;
FEATURES Location/Qualifiers
source 1..3750
/organism="Non-A, non-B hepatitis virus"
/mol_type="unassigned DNA"
/db_xref="taxon:12440"
<1..>3750
/codon_start=1
/product="partial NANBH-PT polyprotein"
/protein_id="CAA02095.1"
/db_xref="GI:1926563"
/translation="MEGVFTGLTHVDHAFLSQTKQAGNPFYLVAVQATVCARAQAP
PSMDQMKCLRLKLTGPTPLVRLGAVQNEVTLTHPTLTFIMACMSADLEVVTST
WLVGVLAALAAVCLTGTGSVIVIGRIILISGRPAIVPREVLVGFDEMEBCASHLPY

```

CDS

```

1..3750
/organism="Non-A, non-B hepatitis virus"
/mol_type="unassigned DNA"
/db_xref="taxon:12440"
<1..>3750
/codon_start=1
/product="partial NANBH-PT polyprotein"
/protein_id="CAA02095.1"
/db_xref="GI:1926563"
/translation="MEGVFTGLTHVDHAFLSQTKQAGNPFYLVAVQATVCARAQAP
PSMDQMKCLRLKLTGPTPLVRLGAVQNEVTLTHPTLTFIMACMSADLEVVTST
WLVGVLAALAAVCLTGTGSVIVIGRIILISGRPAIVPREVLVGFDEMEBCASHLPY

```

```

1EQMQLAEQKQKALGLLTATYQAEAAAPVESKRALETFAKHMNFISG1QYL
AGSLTPENPAIASIMAFPAIVTASVTEPLTOSTLLNLTLOGVNPENSLPSPASAFAG
IAGAAGSISGAKGLVDILAGAGGVAALAFKMSGEMSTEDLNLPAISLPS
LVGVGAKIIRRHVGPBGAVOMNNRILAFSGNNHVSPIHYVESPAARVTOILS
DLITTOULKRIHQHINEDCSPTGSSWLRDWDITCYLADPKTKMLQSKLPRLEGVP
FFSCORGKYGWRGDIMQTTCSGAQITGVHKNQSNMIVGPKTCSNMHGTFFPNAY
TTGCTPSPAPNSALMRVAEEVEYTRGDFHYVTSMTDNKVCQVPAPPEFT
EVDGVRHRAVPACKPLREBVTFOVGNOLVLSGOLCEBEPDAVLTSMITDPSHI
TAETKRLAAGSPSLSSASQASASRATYITODPDAIIEANLIMRHEMG
DITRVSENRKVLIIDSFDLRAEDREVSVPATIRKSKKFPMPAMPADRYNPL
LESMAEDPYVPVHAGCLPPTKPIPIPRPRKTVUTESTVSALAEATKAGSS
EPASVDSGTATAPDPQSDDGASDVESYSMPLEBEPDPLSDGSSVSEAG
EDVCCSMSYVTMGALITPCAAEBSKPIINALSNLHNMHMYATTSRSQKRY
TFDRLOILDHYODVLKEMKAKASTVKAKLISVEBACTLTPHSAKSKFYGAKDVRN
LSSKAIHIRISVMEDLIEDTEPIDTITMAKNEVFCQPEBGRKPARLIVPELIGVR
VCEKRALVDVSTLPOAMGSSYSGVSGRQVRELVNAMSKTQPMGFAYDTGCFDS
VTREKALDIVESITDCCCLAPBARQIHSILRLRYIGPPLNRSGQNGRYRCRASGV
LITSQNTLTCYLKASACRAKLDQDMVLVCGDLVVICBSAGTQEDPAASLRVTEA
MTRVSAPGDPDPPEYDLELITSCSSNSVVAHDSGRKRYVLTBDP"

```

ORIGIN

```

Query Match 99.0%; Score 2022.2; DB 6; Length 3750;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 2030; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

```

```

Qy      1 TGGGAGGCGCTCTTCAACAGGCGCTCACCGAGTGATGCCACTTCTGCTCCCAACAAAG 60
Db      1 TGGGAGGCGCTCTTCAACAGGCGCTCACCGAGTGATGCCACTTCTGCTCCCAACAAAG 60
Qy      61 CAGGAGAGAGACAACTTCCCTTACTGCTGAGGCTAAGCTGCTGCTGAGGCTG 120
Db      61 CAGGAGAGAGACAACTTCCCTTACTGCTGAGGCTAAGCTGCTGCTGAGGCTG 120
Qy      121 CAGGCGCCCATCTCCATCATGAGATGAAATGAGAGTGTCTCATACGCTAAAGCTACT 180
Db      121 CAGGCGCCCATCTCCATCATGAGATGAAATGAGAGTGTCTCATACGCTAAAGCTACT 180
Qy      181 CTGCGGAGGCGCAACACCTTCTGCTGATGAGGCTGAGGAGCGCTCAAAAGAGGTCACCTC 240
Db      181 CTGCGGAGGCGCAACACCTTCTGCTGATGAGGCTGAGGAGCGCTCAAAAGAGGTCACCTC 240
Qy      241 ACACACCCCATTAACCAATTCATCATGAGCATGATGTCAGCCGACTGAGAGTGTCAACG 300
Db      241 ACACACCCCATTAACCAATTCATCATGAGCATGATGTCAGCCGACTGAGAGTGTCAACG 300
Qy      301 AGCACCTGGGTCTGTGGGCGGGGTCTTGTGACGCTCTGCTGCTATTTGCTTGACAACA 360
Db      301 AGCACCTGGGTCTGTGGGCGGGGTCTTGTGACGCTCTGCTGCTATTTGCTTGACAACA 360
Qy      361 GGCAGCGTGTGATGTGAGTATGATCATCTTGTCCGGGCGGCGGCTATTGTTCCGAC 420
Db      361 GGCAGCGTGTGATGTGAGTATGATCATCTTGTCCGGGCGGCGGCTATTGTTCCGAC 420
Qy      421 AGGGAAGTCTCTTACAGAGATTCGATGATGAGAAAGTGGCGCTGCACCTCCCTTAC 480
Db      421 AGGGAAGTCTCTTACAGAGATTCGATGATGAGAAAGTGGCGCTGCACCTCCCTTAC 480
Qy      481 ATCGAGCGAGGATACAGCTGCGCGAGCATTTCAAGCAAAAAGCGCTGGGTGTCTGAG 540
Db      481 ATCGAGCGAGGATACAGCTGCGCGAGCATTTCAAGCAAAAAGCGCTGGGTGTCTGAG 540
Qy      541 ACAGCCACAGCAAGGAGGAGGCGCTGCTCCCGGTGTGAGTCCAAAGTGGGAGACCTT 600
Db      541 ACAGCCACAGCAAGGAGGAGGCGCTGCTCCCGGTGTGAGTCCAAAGTGGGAGACCTT 600
Qy      601 GAGACCTTCTGGGCGAAACACATGTGAATCTTCAACGCGGATACATTAATGACAGGC 660
Db      601 GAGACCTTCTGGGCGAAACACATGTGAATCTTCAACGCGGATACATTAATGACAGGC 660
Qy      661 TTGTCACTCTGCTGAGATTCGCGGATTTGATTAATGATGAGGCTTCAAGGCTCTGTG 720
Db      661 TTGTCACTCTGCTGAGATTCGCGGATTTGATTAATGATGAGGCTTCAAGGCTCTGTG 720

```

| | | | | |
|-----------|------------|--|---|------|
| | Db | 1801 | TGCGAGCTCCCATGCGAGCCCGAACCAGATGTACAGAGCTCACTTCATGCTCACCGAC | 1860 |
| | Oy | 1861 | CCTTCCACATCACAAGAGAAGACGGCTTAAGCGCAGGCTGCACAGGGGTCTCCCCCTCC | 1920 |
| | Db | 1861 | CCCCTCCACATCAACAGAGAGACGGCTTAAGCGCAGGCTGCACAGGGGTCTCCCCCTCC | 1920 |
| | Oy | 1921 | TTGGCCAGGCTCTTGAGCTAGCGCACTTGTCTGCGGCTTCTCGAAGGCCACATACATTACC | 1980 |
| | Db | 1921 | TTGGCCAGGCTCTTGAGCTAGCGCACTTGTCTGCGGCTTCTCGAAGGCCACATACATTACC | 1980 |
| | Oy | 1981 | CAAAATGACTTCCAGACGCTGACCTCATCGAGGCCAACCTCTGTGGCGGCATGAGATG | 2040 |
| | Db | 1981 | CAAAATGACTTCCAGACGCTGACCTCATCGAGGCCAACCTCTGTGGCGGCATGAGATG | 2040 |
| | Oy | 2041 | GGC 2043 | |
| | Db | 2041 | GGC 2043 | |
| | RESULT 8 | | | |
| | HPCKIS2 | | | |
| | LOCUS | | | |
| | DEFINITION | 9410 bp RNA linear VR1 10-FEB-1999 | | |
| | VERSION | Hepatitis C virus (strain HCV-1b, clone HCV-KI-S2), complete genome sequence. | | |
| | KEYWORDS | D50485.1 GI:1030704 polyprotein; core protein; E1; E2; NS2; NS3; NS4A; NS4B; NS5A; NS5B; envelope protein; non-structural protein; interferon-sensitive; interferon-resistant; IFN-resistant; ISDR; interferon sensitivity determining region; HVR; hypervariable region. Hepatitis C virus Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus. | | |
| SOURCE | | | | |
| ORGANISM | | | | |
| REFERENCE | | | | |
| AUTHORS | | | | |
| TITLE | | | | |
| JOURNAL | | | | |
| MEDLINE | | | | |
| PUBMED | | | | |
| REFERENCE | | | | |
| AUTHORS | | | | |
| JOURNAL | | | | |
| REFERENCE | | | | |
| AUTHORS | | | | |
| TITLE | | | | |
| JOURNAL | | | | |
| FEATURES | | | | |
| SOURCE | | | | |
| CDS | | | | |

VDLGCVFLVSQLTFTSPRRHETVQDNCISLYPGHSGRHMANMDMMNSPTTALVV
SGLRLIPOAVMDMVGAGHVLGAGLYASMMNMAKVLIVMLFVGVRGRTVTCGAQ
GHTTORLASLFTFGAQRLOLVNTNGSMHINTANGLNLGTLATGFLYRHNSS
GPPBRVIVGTTIDAFACGMGPIITVAEPPSSDORPCYHAPRPGCIIVASRVCQVVC
PPCNIGGVGNNTLTCTPDCFRKHPRATYTKCSGPMWLRPCWVDPYRLMHPCTVNF
TIFKRMVYGVGEHRLNAAICMWTGERCDLEDRDSELSPLLDSTAEQOILPSCSTTL
PILSTGLIHLHONIVDOYLYGVSAVSVFKMEYIILLFLLDARCAACMMMLL
IAOEAALENLVNVAASVAGHGLFSLFPCAMRYKLPVPAAYAFVYMLL
LILALPRAYAMDRMAASCGGAVFVGLALLSPHYVYVRLALWMLQYFTRBAC
LQWMLPLANVGRRAIILLTICALHBEILFSTIKLVAILGSLMTLQSGITMPFVR
AOGILRACMLVRKVAAGHYVQAFMKLAALTGVYDHLTQDAAHGLMRLAAVE
PVFSDMETKLIITWADTAACGDILGLASVARSKEIFLGPADLESBMGLAPIT
AYSQRTGLGCIITSLTGRCQVEGEVQVSTAGSLTCTGTCVTCVTHVAGSK
TLASGKPIITOMYITVDOLVGMQAPSGARSILPCTGSSDLVLTGRADVYVRGR
DGRSLSPRVSYLKSSGGBPLCPSHAVGIPRAACTRGAVALPIPESMETT
MRSPIYFNTSSPPRPIQTFQVAAHLAFTGSGSKNKPVEYXACQKVLVYNVSAATL
FGVAYMSRAHGVDPNIRTGVRTITTGAPITYSTYKFLADGCGSGADIIILCDECHS
TDSISILGIVLDOAETAGARLVLAATAPGSAVTPHPNIEBALNIGBIPEYK
AIPLETTIKGRHLIFCHSKKCDLAAKSGIGNAVAYRGLDVSVIPASGDVVVA
TDALMTGFDGDSVIDCNTCVTOTVDSLDPTLETITVPODASRSORGRGGRG
RAGIYRVTGRBPSGMDSSVLCQCYAGCAWELTAPETSRLRAYLNTPGLEPWO
DHEFMSVFTGTLTHIDAFISQTOAGDNPRYLVAQATVCARQAAPPSDQMKC
LIRKPLTHGPTPLRLGAVQNEVILTHPIKYIMACMSADLEVITSTVLVGVLA
ALAACTTGSVIVGRIILSGKPAIIPDREVLYPEPMEBCASHLPYIEGMLAB
QFOKALGILQATKQAEAAAVESKNOLETFWAKHMMNFIISGIVYLAGSLTSPN
PALASIMFTASITSPLTTOHTLFINIIGWVAQAOLPPRASAIVGAGIAAAGSI
CLGKLVNDIAGXGAVAGALVAFKVMGMDSTEDIVNLLPAILSPRALVYGVCA
ILRHVYFEGEAVQMMNRLIAPASRGNHSPTHYVESDAARVQIISLITLTLK
RLHWINDECSPTSCGSMRLDWDMLCTVLDPKTQLSRVLRPLGVFPLSCQKQK
GVMRGDIMQITTCPCGAQITGHVKNKSMIVGPKCSNTHGTPFINAYTGPCTSP
APNYSRALMRVAAEVEYTRVGDHVTGMDNLKCPQCVPAEPFEEGSEVAVCCMS
YTABSKPLRDEFTVOGLNOVVGSLPCEBEPRVAVLTSMLTDPSHITAAARBL
ARGSPSLASSASQSLAPSKATCTTRHSDPADLILANILMRBENGNTIRVSEN
KVLILDSFEPFAEEDEREVSIPAEILKSRKFPAMPMAWDPNPLIESWKPDY
VPPVHGPCLPPTKAPPIPPRRKRTVLTSTVSALAEATKTFGSESSAASDGT
ATAPDQSNDDAGSDVESYSSMPLEBEPDPLDSGMSGTVEEAGEEDVAVCCMS
TYTGALITPCAEBSKLPINALNSLRHNMVYATTSRSQKQKTFPRLDVL
DHYRVLKEMKASTVAKRLILVEACKLTPHSAKRFSGAGADVNLSTLNI
RSVMKDLEDTETPTDITIMAKNEVFCQPEKGRKPARLIVPEPLGRVCEKMLYD
VSTLPOAVMGSYGFQYSPQGRVEFLVNAKSKCPMGFAIDTCEPSTVESIRV
BESIIQCCDLAPEARQIRSLTERLYIGGLPLNSKQNGYRCHASGLTATSCNIR
TCYLKATACRAKACQDCTMLVAGDVLVICSAGSTPEASRLAETEMTYSAPG
DPOEYVLEILITSCSNVSAHDASGRVYVLTDPPTPLARAMEBTAHYPVSWL
GNITMAYAPTLAABMILMTHFBSIILAOQLKALDCQIYGCXYSIEPDIPTIOLRH
GLSAPLSHSYSPGEINRVAASCLRKGLVPLRWRRHARSVRKLLISQGRRAICGRTL
FNMAVRTKLPLIPAAQDLDSGWVAGVSGSDIYHVSRRARPRWFMLLLLSGV
GIYLLPNR"

mat_peptide 330..902
/product="core protein"
903..1478
/product="E1"
1479..2756
/product="E2"
2757..3407
/product="E2"
3408..5300
/product="NS3"
5301..5462
/product="NS4A"
5463..6245
/product="NS4B"
6246..7586
/product="NS5A"
7587..9359
/product="NS5B"
1479..1559
/note="hypervariable region (HVR)"
6954..7073
/note="interferon sensitivity determining region (ISDR)"

misc_feature
misc_feature
misc_feature

Query Match 88.3%; Score 1804.6; DB 14; Length 9410;

Best Local Similarity 92.7%; Pred. No. 0; Matches 1894; Conservative 0; Mismatches 149; Indels 0; Gaps 0;

1 TGGGAGGGGCTTTCACAGGCTCAACCAAGGATGCGCACTTCCTGACCAAAAG 60
5001 TGGGAGGGGCTTTCACAGGCTCAACCAAGGATGCGCACTTCCTGACCAAG 5060

61 CAGGAGAGAGCAACTTCCCTCACTGCTGAGGATGAGGCTTCTGCTGAGGCT 120
5061 CAGGAGAGAGCAACTTCCCTCACTGCTGAGGATGAGGCTTCTGCTGAGGCT 5120

121 CAGGCTCCCACTTCATCATGAGGATGAGGATGAGGATGAGGATGAGGATGAG 180
5121 CAGGCTCCCACTTCATCATGAGGATGAGGATGAGGATGAGGATGAGGATGAG 5180

181 CAGGCTCCCACTTCATCATGAGGATGAGGATGAGGATGAGGATGAGGATGAG 240
5181 CAGGCTCCCACTTCATCATGAGGATGAGGATGAGGATGAGGATGAGGATGAG 5240

241 ACACACCCCATTAACCAATTCAATGAGGATGAGGATGAGGATGAGGATGAGG 300
5241 ACACACCCCATTAACCAATTCAATGAGGATGAGGATGAGGATGAGGATGAGG 5300

301 AGCAGCTGAGGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 360
5301 AGCAGCTGAGGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5360

361 GCGAGCGGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 420
5361 GCGAGCGGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5420

421 AGGGAAGTCTCTACCAAGAGTTCATGAGGATGAGGAGGAGGAGGAGGAGG 480
5421 AGGGAAGTCTCTACCAAGAGTTCATGAGGATGAGGAGGAGGAGGAGGAGG 5480

481 ATCGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 540
5481 ATCGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5540

541 ACGGACCCCAAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 600
5541 ACGGACCCCAAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5600

601 GAGACCTTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 660
5601 GAGACCTTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5660

661 TTGTCCACTCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 720
5661 TTGTCCACTCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5720

721 ACTAGCCGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 780
5721 ACTAGCCGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5780

781 GCGCAACTGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 840
5781 GCGCAACTGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5840

841 GCTGTGAGCAGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 900
5841 GCTGTGAGCAGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5900

901 GAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 960
5901 GAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5960

961 GACCTGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1020
5961 GACCTGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6020

1021 TGGCAGGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1080

Db 6021 TCGCAGCAATATCTGCGTGGCATGTGGCCCGGAGAGGGGGCTGTGACATGTGATGAC 6080
Qy 1081 CGGCTGATAGCGTTGCGCTCGCGGGGTAAACATGTTTTCCCGCAGCAGCATATGTGCAGAG 1140
Db 6081 CGGCTGATAGCGTTGCGCTCGCGGGGTAAACAGTCTCCCGCAGCAGCATATGTGCTGAG 6140
Qy 1141 AGCGAGCGCGCAGCAGTGTCACTAGATCTTCTCCGACTTACATCACTCACTGCTTTG 1200
Db 6141 AGCGAGCGCGCAGCAGTGTCACTAGATCTTCTCCGACTTACATCACTCACTGCTTTG 6200
Qy 1201 AAGAGGCTCCACAGTGAATTAACAGAGAGCTGTCCAGCGCCGTGCTCGGGCTCGGGGCTA 1260
Db 6201 AAGAGGCTCCACAGTGAATTAACAGAGAGCTGTCCAGCGCATGCTCGGGCTCGGGGCTA 6260
Qy 1261 AGGAGTGTGGGAGCTGATATGACAGTTTTGGCTGACTTTCAGACCTGAGCTGACGTC 1320
Db 6261 AGGAGTGTGGGAGCTGATATGACAGTTTTGGCTGACTTTCAGACCTGAGCTGACGTC 6320
Qy 1321 AAGCTCTGCGCGCATTTACCGGGAGTCCCTTTTCTCAATGCCAAGCTGGGTACAGGGG 1380
Db 6321 AAGGTCCTGCGCGCATTTACCGGGAGTCCCTTTTCTCAATGTCAAGCTGGGTACAGGGG 6380
Qy 1381 GCTTGGCGGGGAGAGGAGCATATGAGACCACTGCTCATATGAGAGCAGATCAACCGGA 1440
Db 6381 GCTTGGCGGGGAGAGGAGCATATGAGACCACTGCTCATATGAGAGCAGATCAACCGGA 6440
Qy 1441 CATGTCAAAAACGTTTCCATAGAGATCGTTGGGGCTTAAGACCTGTATGATTAACATGTGGAT 1500
Db 6441 CATGTCAAAAACGTTTCCATAGAGATCGTTGGGGCTTAAGACCTGTATGATTAACATGTGGAT 6500
Qy 1501 GGAACATTTCCCATCAAGCAGATACACACGGGGCCCTGACAGCCCTCCCGACGCCAAGC 1560
Db 6501 GGAACATTTCCCATCAAGCAGATACACACGGGGCCCTGACAGCCCTCCCGACGCCAAGC 6560
Qy 1561 TATTCAGGGGGCTGTGGCGGGTGGCTGTGAGAGATATGTGAGAGTTAACGGGGTGGGG 1620
Db 6561 TATTCAGGGGGCTGTGGCGGGTGGCTGTGAGAGATATGTGAGAGTTAACGGGGTGGGG 6620
Qy 1621 GATTTCCATACGATACAGAGCATGACCACTGACAAAGTAAATGGCCCGTCCAGAGTTCCA 1680
Db 6621 GATTTCCATACGATACAGAGCATGACCACTGACAAAGTAAATGGCCCGTCCAGAGTTCCA 6680
Qy 1681 GCGCCCGAATTTTTCACAGAGTGGAGTGGCGGCTGACAGATGACGCTTCGGCGTGC 1740
Db 6681 GCGCCCGAATTTTTCACAGAGTGGAGTGGCGGCTGACAGATGACGCTTCGGCGTGC 6740
Qy 1741 AAACCTCTCTACGGGAGAGGATCAATTCACAGTCCAGGTCGAGCTCAACCAATACCTGGTTGG 1800
Db 6741 AAACCTCTCTACGGGAGAGGATCAATTCACAGTCCAGGTCGAGCTCAACCAATACCTGGTTGG 6800
Qy 1801 TCGCAGCTCCATGCGAGCCCGAACCGAGTATGAGCAGTGTCTCATGCTCAACCGAC 1860
Db 6801 TCGCAGCTCCATGCGAGCCCGAACCGAGTATGAGCAGTGTCTCATGCTCAACCGAC 6860
Qy 1861 CCCTCCCATATCAACAGCAGAGACGGCTAAGGCGAGGCTGAGCGAGGGGTCTTCCCTCC 1920
Db 6861 CCCTCCCATATCAACAGCAGAGACGGCTAAGGCGAGGCTGAGCGAGGGGTCTTCCCTCC 6920
Qy 1921 TTGGCAGGCTCTTACGCTAGCGAGTGTGTGGCGCTTCTCGAAGGCGCAGATCAATTAAC 1980
Db 6921 TTGGCAGGCTCTTACGCTAGCGAGTGTGTGGCGCTTCTCGAAGGCGCAGATCAATTAAC 6980
Qy 1981 CAAATGACTTCCAGACGCTGACCTCATGAGGCGCAACCTCTGTGGCGGATAGATG 2040
Db 6981 CGTCAAGACTCCCGCAGATGCCAGCTCATGAGGCGCAACCTCTGTGGCGGATAGATG 7040
Qy 2041 GGC 2043
Db 7041 GGC 7043

RESULT 9
AB049088

LOCUS AB049088 9616 bp RNA linear VRL 22-AUG-2002
DEFINITION Hepatitis C virus genomic RNA, complete genome, isolate:HCVT094.
ACCESSION AB049088
VERSION AB049088.1 GI:11559442
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
PUBMED
AUTHORS
REFERENCE
JOURNAL
TITLE
AUTHORS
FEATURES
source
1. .9616
/organism="Hepatitis C virus"
/mol type="genomic RNA"
/isolate="HCVT094"
/specific host="human"
/db_xref="taxon:11103"
/tissue type="serum"
/note="HCV-genotype: 1b"
1. .341
/note="complete"
342. .9374
/codon_start=1
/product="polyprotein"
/protein_id="BAB1.8801.1"
/db_xref="GI:11559443"
/translation="MSTNPKPORKTKNTNRRPODVKEPGGQIVGVYLLPRGRPL
GVRAKRTSERSQRRGRROPVVKARPPGRWADPGVWPLVGMGMMAALMSPRG
SRPMGPTDPRRRSRNKGVIIDITLTGCPADLMGTVPLVGAFLGGAAPLACVVLTD
SVNATGNTLPGCSFSLPLALSLCTLPASVEVHNSGTHVINDCSNASTVYBAD
MIMHAPGVPCPRDNASRCWVALPTPLAENASVPVTTIRHVDLVGAAPFCAMV
VGDGCVFLISQLETFSPRHETVQDNCISIVGHVTHGMADMMAMPTALVV
SOLRIPOAVMDVAGAHVGLAGLAVYSMAQNAKYLIVMLFAGVDGNHVSAGVQ
GRVYSILANTFSLGASOKIQIINNGSMHINRTALNCDLSIQTFIALPAFKRNAS
GCEPRMASCRSIDPFGDMGPITYAQSBNLDQRYCHRYAQQGCIYVASQVCEPYC
FTSPVAVGTTDSSGVPPTYSGNENETDVLNLNTRPQGNFGCTMANSIGFTKCGG
PPCNIIGVGNNTLTCPTDCEPKHBEATYTKGSGPWLPRCLVDPYRLMRYPCVNF
TIPIKRMVGVGEHRLHAACMTGERDLEDRSELSPLLSTTTEWQIIPGFTTL
PALSTGLJHQNIVDOYVLIGISAVSFPAIKMEYVLLPLADARVCACLMMMLL
IAQDAALENIVYNAASVAGAHFLSTVPRCAAWYIKGRLVGVAVYALYGPILL
LILALPREPAMDRBMASCGAVFVGLALITLSPYTKVLAELIMLQIYIAAEAH
LQWVPLPLNVRGSDAIILITCAHPELIPITLKLALIGLPLVFGQIIRVYFVR
AQGILRYOMLVKRYAGHYVQMALMKLALGTGVYHLPLRDMANAGLDLAAVE
PVSFDMETKLIITWAGDPAACGDIISGLPVSARGKEILLPADSLBGRRLAPIT
AYSTQGLIGCIITSLGRDKNOVEGVVSTATOSFLASCNVGCVMTVYHAGSG
TLAQRKEPITQMTNNODLVGMNAPGABSMPTCGSSDLYLVTHADIVPYRRG
DGRSLSLPREVSTYKSSGGPLCPGSHVVGITRAAVCTRGAKANDPVESEMETT
MRSPVFTDNSSPPAVPQTFVYLAHAPFGSGSKTKVPAVYAQIKVILVPSVATL
GFGVMSKASHSTDNIRITGRTITITGAIPITVSTGKFLADGCGGAYDIIICDCHS
TDSITLIGITVLPQAEATAGARLVLVLAATPFGSVTPVHPNIEVALSNTGEIPYCK
AIPLEAIKGERHLFCSSKRCDBLAKLSGLINATAVYRGLDVSIVIPSGDVVVA
TDAIMTGFTGDPDSVIDNCTVOTVDPDSLDPTTETITTYTPDAVSTQRRGRTGRG
RRIYREVTPEGRSPSGMDSVLCECDAGCAMPLEPAETSVLRATLNPBGIPVCO
DHLBFWESVFTGLTIDAHFLSQTKQADNCPVLAQVATVCARQAPSPMDQWKC
LIRLKPILHGPTPLLYRGAVHNEVTLTHPTKXYIMCMSADLEAVTSTVLGVGLA
ALAAVCLTGGSVIVGRIILSGRAVVIDPRLVYREPEMECSHLYPIYIQGMLAE
OFKOKALGLOTATTOAEAAAPVSESKRALEAFQAKMNNPISGIVYLAGISTLPGN
PALSMAFTASTISPLTOSTLFLNLGCVNVAQALAPPSAASAPVAGAGIAGATGSI
GLGRVLVDIILAGYAGVALVAFKWSGEMFTBEDVNLPLILIPALVVGIVCAA

Db 6693 ACCCCGAAATCTTCAACGAGGTGATGAGTGGCGCTGCAAGGTACGCTCCGCGTGC 6752
 Qy 1741 AAACCTTCCTCAGGAGAGGTGCAATTCAGAGTCGAGGCTAACCAATACCTGGTGGG 1800
 Db 6793 AAACCTTCCTCAGGAGAGGTGCAATTCAGAGTCGAGGCTAACCAATACCTGGTGGG 6812
 Qy 1801 TCGCAGCTCCATGCGAGAGCCGCAACCGAGTGTAGCAGTGTCTCATTGCTCAACGAC 1860
 Db 6813 TCACAGCTCCATGCGAGAGCCGCAACCGAGTGTAGCAGTGTCTCATTGCTCAACGAC 6872
 Qy 1861 CCCTCCCATCAGCAGCAGAGCGCTTAAGGCGCAGGCTGCGCAGGGGGTCTCCCCCTCC 1920
 Db 6873 CCCTCCCATCAGCAGCAGAGCGCTTAAGGCGCAGGCTGCGCAGGGGGTCTCCCCCTCC 6932
 Qy 1921 TTGGCAGCTCTTCAAGCAGTGTCTGCGCCTTCTCGAAGCGCAGTACATTAAC 1980
 Db 6933 TTGGCAGCTCTTCAAGCAGTGTCTGCGCCTTCTCGAAGCGCAGTACATTAAC 6992
 Qy 1981 CAAATGACTTCCAGAGCGCTGACCTCATCGAGGCGCAACCTCTGCGCGCATGAGATG 2040
 Db 6993 CATGATGACTTCCAGAGCGCTGACCTCATCGAGGCGCAACCTCTCTGTGTGCGAGGATG 7052
 Qy 2041 GGC 2043
 Db 7053 GGC 7055

RESULT 10 AF333324

LOCUS AF333324 9587 bp mRNA linear VRL 18-DEC-2002
 DEFINITION Hepatitis C virus type 1b polyprotein mRNA, complete cds.
 ACCESSION AF333324
 VERSION AF333324.1 GI:12831192
 KEYWORDS Hepatitis C virus type 1b
 SOURCE Hepatitis C virus type 1b
 ORGANISM Hepatitis C virus positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.
 REFERENCE 1 (bases 1 to 9587)
 AUTHORS Thomson, M., Nascimbeni, M., Gonzales, S., Murthy, K.K., Rehmann, B. and Liang, J.J.
 TITLE Emergence of a distinct pattern of viral mutations in chimpanzees infected with a homogeneous inoculum of hepatitis C virus
 JOURNAL Gastroenterology 121 (5), 1226-1233 (2001)
 PUBMED 21534507
 MEDLINE 11677216
 AUTHORS Thomson, M. and Liang, J.
 TITLE Direct Submission
 JOURNAL Submitted (03-JAN-2001) LDS, NIDDK/NIH, 10 Center Drive, Bethesda, MD 20892, USA

FEATURES source

1..9587 location/Qualifiers
 /organism="Hepatitis C virus type 1b"
 /mol_type="mRNA"
 /strain="chimera of HCV-BK, HCV-J, HCV-H"
 /serotype="1b"
 /specific_host="chimpanzee"
 /db_xref="taxon:31647"
 /clone="HCV-CG1b"
 1..341
 1..271
 1..271
 /note="derived from BK strain; similar to sequence deposited in GenBank Accession Number M58335"
 272..3030
 /note="derived from J strain; similar to sequence deposited in GenBank Accession Number D90208"
 342..9374
 /codon_start=1
 /product="polyprotein"
 /protein_id="AAK08509.1"
 /db_xref="GI:12831193"
 misc_feature
 misc_feature
 CDS

/translation="MSTPRKQRKTRNRPDVRFPGGQVGVGVYLLPRGRPL
 GVARIRKTSRQPRGRROPFKARPEGRTWAPGPYPLVGNBGMQAMVLSPRG
 SRPMGPDPERRSRNLKVIDLTLCGPDAIDMGYIPIVGAELGAGARLGAIVLS
 GNVATGILPGCSFSEIILALISCLTIPASAYEVRANYSIHVINDGNSIIVEAD
 MIMHPGVCPREBSNFSRCWALPTLPAALPTLPTLIRYVLDLGAALCSAMY
 VGLDLSGVFLVSLPTFSPPRYETVODQNSIYGHGSHGMDMMNSPTLAVY
 SOLLRTPQAVDVMAGAKVILAGLAIYSNKGMAKILIVMLFRAGVDGHTHGRV
 ASSTQSLVSLSGSPQKTLQVNTNGSHIRKTLNCLNDSIQTGILALFVHRFMS
 GCPBRMASCRPIDKFAQGMGIITHVNPISQORCAWYAPQGIIPASQGVPCV
 FTPRNVVGTDRSGVPTPYSWGNETVDLLNTRPPDGMFGCTGNMSTFTTCGG
 PCNIGVGNVNTLCPTDCFRKPEATYKSGSPMLPTPLCLVDPYPLMYPCITNF
 TIFPRVNGVEHRIINAACNMTGGEEDDLEDRSELSPILSTTEMQVPCSTTL
 PALSTGLHLMQNVVVOYLIVGVSVYVSVIKREYTLLELLADAPVCAIIMMLL
 IAOEATLENIIVNLAASVAGAGLISLTFVFCAMYTIGRLVGAALVAGVPL
 LLLALPRVAMDERMASCGAVFVGLVLTLPYKVFARLIMLQYITREAH
 LQVWPLPLVAGRDAILLILCAHPELIPITLLALILPMLVLOAGIRVFPVR
 AOGHILACMLVRKAGGHVYOMAFMKALGTGYIYHMLPLRDMHAGLIDVAVE
 PIVRSDETETKILMGADPAACGDIILGIPVARBGKELILGADSLBERGRLAPIT
 AYSQTRGLCCITITSLGRNRYBEGVYVSTRATQSEFLATCVNVCWYVYHAGASK
 TLAPKGPITQMYTNVDLVLGWPAPPEARLTFCTGSSDLVTRHADYIPRRRG
 DSRGSLSPREVSYLKSSGGPLCPSSHAAGIIPAACTGVAKAAVDVPEVSEMTT
 MRSVFTDNSPPPAVPOSFQVAHLHAPGSGSKTKVPAVYAOQKXVILNPSVALT
 GEGAYMSKAGHIDINIRGVRTITGAIVTYSTYKFLADGCGGAGADIIICEHS
 TDSPTIIGITVLDQAEAGARLVVLAATPPGAVTPHPNIEVLAISNTEIPEYK
 AIPETIRGGRHLIFCHSKKCDLAKLSGLGINAVAYRGLDVSVPITGIVVVA
 TDLMTGYTGDPSVDNCTVQTVQVPSLDPTITLTITVPODASRQRTGRG
 RRGIVREVTPEGRSGMFDSSVLCEDYDAGAWELPAETSVRLAVLNTPLGLVCO
 DHEFWESVFTGLTHIDAHFLSOTKQADNPYVAVYQATCARAQAAPPMDQWKC
 LIRKPTLHGPPLTYRGAVONETVTLHPITKTYIMACMSADLEVSTVTLGVGLA
 ALAAVCLTGVSVIVGRIISGRPAIDRELLVLOEPDMECEATHLPIEIQMOLAE
 OFKOKALGLTOTATKQEAALPVYBSKRALETFRMAKMMNPIGSIQVLAELSTLPG
 PAISLMAFTASTISPLTQSTLLFNILGAVNAQALAPSAHAFVAGIAGAVLSI
 GLGVLVLDILAGYAGVAGALVARKVMSGEMSTEDLVNLLPALSLPALVGVCAA
 ILRRHVGEBAVOMNRLIAPASRGHVSPTHVPEBDAALAVYQILSLITQLK
 RLHOMINDECTPCSGSLRDVMDMICYLDFKTLQSLKLPOLGVPPFSQGRYK
 GWRMGDQMTTCPCGAGQITGSHVNGSMRTVGPCTCAHTMGTPFINAYTGPCTPSP
 APNYSRALMRVAEEYEVTVGDPHYTGTGTTDNVKKPCQVPAPEPSEBVDGRLHR
 YAPACRPLIREVTFVGLNQLVLSQSLCEPDEVAALVIMLTDPSHTIETARL
 ARGSPPLASSASQSLASPLKATCTHVPBDDLLEANLMLGEMGNTITRVESN
 KVVVLSDFDPLRAEDEREVSIPAEILKSKKPPAPAMPINARPDYNPILSSWQPDY
 VPPVHGCEPLPIKAPPIPPRRKRTVLTSSYSALAEATLTFGSSSSASVDSGT
 ATALPDQASDDGSDGDSYSMSWPLGEGGDDPLSDGMSSTVSEBSGVSQMS
 YTMGALITTPCAABESKIPINALNSLIRHNNYATTSRSAGLRQKRYTDRDLQVLD
 DHYVDVLEKEMAKASTYVAKLISYBEAKLTPHSKASFEYGADVNLSKAVNHI
 HSWVKQDLEEDTPTIDTITMAKNEVFCQPPKGRKPRLLVFPDGRAVCEKALYD
 VSTLPQVWMSYGFQYSPQARVEFLVNTWKSXKPNPSPGYDTECPSTYENDIRV
 BESLYOCCDLAPBAROAKSLTERLYIGGPLTNSGONGRGRCAASGLVLSCENTL
 TCYKASAAACAAKTDCTMLVNGDILVTCESAGTSCGDSASLRYFTAMRYSAAPG
 DPQPEYDLELITSCSSNVSVAHDSGRKRYITRTDPTPLARAWETARHTPVNWSL
 GNIMYAPTLWARMILMTHFFSILAOQLKALDCQIYGACYSIEPLDLPQILRLH
 GLSAFSLHSYSPGIEINRVASCLRKLGVPLEVWRHARSVBARLSOGGRATYCKYL
 FNMVAVTKLKLTPIPASQDLDSGFVAGSGDGIYHSLSPARPRFWMLCLLLSVGV
 GIVLLPNR"
 3031..9404
 /note="derived from BK strain; similar to sequence deposited in GenBank Accession Number M58335"
 9375..9587
 3'UTR
 misc_feature
 /note="derived from HCV-1a strain H; similar to sequence deposited in GenBank Accession Number AF009070"
 ORIGIN
 Query Match 88.2%; Score 1801.4; DB 14; Length 9587;
 Best Local Similarity 92.6%; Pred. No. 0;
 Matches 1892; Conservative 0; Mismatches 151; Indels 0; Gaps 0;
 Qy 1 TGGAGGCGCTTTCACAGGCTTCAACCACTGATGCCACTTCTGTCCCAACAAG 60
 Db 5013 TGGAGAGTGTCTTCAACAGGCTTCAACCACTGATGACACTTCTTCCAGACAG 5072
 Qy 61 CAGGAGAGAGCAACTTCCCTTACTGTGGGTGCGAGGCTACGTGTGGGCTAGGGCC 120
 Db 5073 CAGGAGAGAGCAACTTCCCTTACTGTGATACCAAGCAAGGTGTGGCGGCT 5132

AUTHORS Nagayama, K., Kurosaki, M., Enomoto, N., Maekawa, S. Y., Miyasaka, Y.,
Tazawa, J., Izumi, N., Marumo, F. and Sato, C.
TITLE Time-related changes in full-length hepatitis C virus sequences and
hepatitis activity
JOURNALT Hepatology 263 (1), 244-253 (1999)
MEDLINE 20013325
PUBMED 10544098
REFERENCE 2 (bases 1 to 9379)
AUTHORS Nagayama, K., Kurosaki, M., Enomoto, N., Maekawa, S., Miyasaka, Y.,
Sakamoto, N., Fukuma, T., Tazawa, J., Izumi, N., Marumo, F. and Sato, C.
TITLE Direct Submission
JOURNAL Submitted (06-JUN-1999) Second Department of Internal Medicine,
Tokyo Medical and Dental University, 1-5-45 Yushima Bunkyo-ku,
Tokyo 113-8519, Japan
FEATURES
source
Location/Qualifiers
1..9379
/organism="Hepatitis C virus"
/mol_type="genomic RNA"
/strain="MD8-2"
/db_xref="taxon:11103"
/note="from a patient with persistently high serum ALT
values; patient 8, point 2
genotype: 1b"
330..9362
/codon_start=1
/product="polyprotein"
/protein_id="AAD56195.1"
/db_xref="gi:5918959"
/translation="MSTNPQRKTKRNTNRPDVRFPGGGQIVGVLLPRRPRRL
GVRAKRTKSERQPRGRQPIPKARPRRGTAAQVQVWPLVGNMGMAAGLLSPG
SRPVGNDPRKSRNLKVIDLTCTGPDLMGVIPLVAGVAGARLAAHGVRLAD
SVNATGM,PGCSFSIPLIALSLCLTIPASAVEVNVGVHVDLNSNSIVVAAID
IIMHPCGTCVAREGINSRCWALTPDAARNISVPTTIRHVDLIVGTAFCSAMY
VGDLCGSEIVLSOLFTFSRPRHETVQDCNSIYPGHVAGRAMMMNWSPTALVV
SOLRIPAVVDMVAGVAGVAGIAGLAAYSMGNMAKVIIMVLFPAGDTHVNGAO
AHSATLMSFTVGPSSOKIQLVNTNGSHIRNTALNCDSIPLTGLAALFTHRNAS
GCEPMASCRPLDITPDQGMPLTAAEPNSPDPRPYCMHPCGIEASRKGCGVYC
FTSPVVVGTIDRFVPTVPMNGENETDVLNTRPPGCMWPGCTMNSGTGKTCG
PPCNIAGAGNTLCTPTDCFRKPRATYKCGSGMPLPRCLVDYPLIMHPCVNF
TIFKRYVGVGEHRLSAACNWTGERGDLERDSLEPLSTEMOCLMCSPTTL
PALSTGLIHRNIVDVOYLGVGSAVSIARMEVYLLPLILADAVOCICMMMLL
IAQEAALENVLVINAASVAGHGLISPLVFCAAMYIKGRVPGAAYALVYRWELL
LLALPRPVAAMDREMAASCGAVFGLAFLTSLPYKVFRLIMVQIPTREBAH
LQWVPLNVRGRDAIILMCVHBEIIFDITKLLALFGLMVLQGIPTREPVFVR
AHGLIRACMLVAKVAGHYVMAFVKLAALGTYYDHLAPLOHMAHSGLRDLAAVE
PVFSDMETKIIITGADTAACGDIILGLPVASRGRREIILGADSLGEGMFLAIP
AYSOOTRGLGCIITSLTGRDKNOYGEVQVSTAGSLAICINVCMTVHVGSGSK
TLAAGRPITQTYTVDODIVQMAPEKARSMTPTCTGSSDIYVITRHADYPIVRRG
DGRGLSPRPVSYLKSGSGGPLCPSGHAAGI FRAAVCTRGVAKAVAFPIVESMET
MNSPVFTNNSPPAVQTFQVAAHLAFTSGSKSTVPAAYAAQGVVLNFSVATL
FGVAMSRAVGTDPNI RTGIRITITGAPITVSTGYKFLADGCGGAAVDIILCDECHS
TDSITLIGITVLDQAEVAGVAVLATTPGSAVTVPHPNIEEVALNSTGSIPIYKQ
AIPRITVGGRHILFCHSKKCDLAAKLSAIGLNAVYVRGLDVSIVIPISGDVVVA
TDALMTGTTGDDSVI DCTCTCTVTVDSLDPTTITETTVQDVA SQRGRGRG
REGIYRFVTPGRPSGMEFSSVLCECYDGCAMYEILPELTSIRAKYITNGLDQVC
DHLBEWSEVFTGLTHIDAFLEQTCQAQGNFVLAQVLETSVARQAPPSDDWKC
LIRLKPCTHGPPLPLRGLAGVQNEVTLITPKIFIMACGASLEVTSTWLVGCVLA
ALAAVCLTGSVIVIGRIILSGPAVILPREVLYOEBMECAALVPIIEGMOALB
OPKOKALGILQTAQKQAEAAAVFVSEKRALETPAKHMMNLSIGIYIAGLSTLPGN
PAISLMAFTASTITSPITQTOTIMRNILIGGAVTAQALAPSAAYVAGIAGINAVGSI
GAGKVLVDILAGVAGVAGALVAFVMSGEMSTEDLINLPAIILSPGLVGVCAAL
ILHRVAGREGAVQMMNRLIAPASGNHVSPTHYVESPAARVQIISLITVGLK
RLHQMINEDESPGSGSMRLQVDMWICVLDPTKMLGKIPRPAJGFLSCQKQY
GWRGDMQIOTPCGAOITGVHKNKRIIVGKPTCSNWHGCTPILNATVTTGCTPSP
APVSRALMRVAAEEVYVTRVGYVYMGTYMTQNDVNCQCPAPAEPEFNDVGLHR
YAPACKPLIREVITFOVGLNQYLVGSQLPCEPEPVAVITSLMTPSHITATAKRL
AAGSPSLASSASQLSAPSLKATCTTHSDADLILENLMRDEMGNTIRVSEN
KVIIIDSPDLRAEBEREVSAAIRSRKFPALPWAPNDPNPLBEMKPDY
VPPVHVGCLPTKAPIPPRKRITVLTESVSALELATITFGSGSGSAVDSGT
ATAPDQASDDGDKESDYESGSMPLBEGPDPDLSDSMSLTVEBESDVVCSMS
YTTGALITTPCAESKLPINPLNSLLRHHMNYATTSRASLQOKYTPRLQVLD
DHYRVLLKEMKASTVAKLLSIEACKLTPHSARSFGVADVSVLSKAVNHI
RSWBDLLEDTEPTIDTITMAKSEVFCVQPEKGRPARLIVFPDLGVAVCEKMALYD

ORIGIN
Query Match
Beet Local Similarity 92.5%; Pred. No. 0;
Matches 1890; Conservative 0; Mismatches 153; Indels 0; Gaps 0;
88.0%; Score 1798.2; DB 14; Length 9379;
1 TGGAGGGCGTCTTCACAGGCTCAACCCAGGTGATGCCACTTCTGTCCAAACAAG 60
5001 TGGAGAGTGTCTTACAGGCTCAACCAATAGATGCCACTTCTGTCCAAACTAAG 5060
QY 61 CAGGACAGACAACTTCCCTTACTGTGGCGATACAGGCTACTGTGGCTTAAAGGCC 120
DB 5061 CAGGACAGACAACTTCCCTTACTGTGGCGATACAGGCTACTGTGGCTTAAAGGCC 5120
QY 121 CAGGCCCACTCCATCATGAGGATCAATGTGGAAGTGTCTCATAGGCTAAAGCTACT 180
DB 5121 CAGGCTCCACTCCATCATGAGGATCAATGTGGAAGTGTCTCATAGGCTAAAGCTACT 5180
QY 181 CTGCGCGGGCCAAACCCCTTGTGTATAGCTGGAGCGGTCCAAACAGAGTCACTC 240
DB 5181 CTGCACGGGCCAAACCCCTTGTGTATAGGCTAGGAGCGGTCCAAATGAGGTCACTC 5240
QY 241 ACACACCCCATTAACCAATTATCATGTGCAATGATCATGACCGACCTGGAAGTCTGCA 300
DB 5241 ACACACCCCATTAACCAATTATCATGTGCAATGATCATGACCGACCTGGAAGTCTGCA 5300
QY 301 AGCACTGGAGTGTGTGGCGGGGCTCTTGCACTGTGCTGATTTGTTGACAACA 360
DB 5301 AGCACTGGAGTGTGTGGCGGGGCTCTTGCACTGTGCTGATTTGCTGACAACA 5360
QY 361 GGCAGCGTGTATTTGTGGGATGATCATCTTGTCCGGCGGGCGGTATTTGTTCCGAC 420
DB 5361 GGCAGCGTGTATTTGTGGGATGATCATCTTGTCCGGCGGGCGGTATTTGTTCCGAC 5420
QY 421 AGGAAAGTCTCTACAGAGATGTGATGAGATGGAAGTGGCGTCCACACTCCCTTAC 480
DB 5421 AGGAAAGTCTCTACAGAGATGTGATGAGATGGAAGTGGCGTCCACACTCCCTTAC 5480
QY 481 ATCGACAGGGAATCAGCTCGCCAGGACGTTCAAGCAAAAAGCGCTGGGTTGTGAG 540
DB 5481 ATCGACAGGGAATCAGCTCGCCAGGACGTTCAAGCAAAAAGCGCTGGGTTGTGAG 5540
QY 541 ACAGCCACCAAGCAAGCGAGCGCTCTCCCGTGTGTGAGTCCAAGTGCAGACCTT 600
DB 5541 ACAGCCACCAAGCAAGCGAGCGCTCTCCCGTGTGTGAGTCCAAGTGCAGACCTT 600
QY 601 GAGACCTTCTGGGCGAAACATGTGGAATTCATACAGCGGATACAGTACTTACAGAGC 660
DB 5601 GAGACCTTCTGGGCGAAACATGTGGAATTCATACAGCGGATACAGTACTTACAGAGC 6660
QY 661 TTGTCACCTGCGTGGGAATCCGCGATGTCATGATGATGATGATGATGATGATGATG 720
DB 5661 TTGTCACCTGCGTGGGAATCCGCGATGTCATGATGATGATGATGATGATGATGATG 720
QY 721 ACTAGCCCGCTACCAACCAATCTACCTCTGTCTTAAATCTCTGGGGGATGGGTAGCC 780
DB 5721 ACTAGCCCGCTACCAACCAATCTACCTCTGTCTTAAATCTCTGGGGGATGGGTAGCC 780
QY 781 GCCCACTGCTCCCCCAAGTCTCTTCAAGTCTTCAAGTCTTCAAGTCTTCAAGTCT 840
DB 5781 GCCCACTGCTCCCCCAAGTCTCTTCAAGTCTTCAAGTCTTCAAGTCTTCAAGTCT 840
QY 841 GCTGTGACAGATGAGCTTGGGGAAGTCTTGTGACATCTTGGCGGGCTATGAGACA 900
DB 5841 GCTGTGACAGATGAGCTTGGGGAAGTCTTGTGACATCTTGGCGGGCTATGAGACA 900

| | | | |
|----|------|---|------|
| QY | 901 | GGAGTGGAGGCGCGCTGTGTGGCCCTTAAAGCTAAGACGGGCGAAATGCCCTCACCGAG | 960 |
| Db | 5901 | GGGGTGGCTGGGCACTGTGGCTTTTAAAGTCATGACGGGAGATGCCCTCCACCGAG | 5960 |
| QY | 961 | GACCTGTGTAACCTTACTCCCTTGCCATCCTCTCTCCTGTGGCCCTGGTGTGGGGTCTGTG | 1020 |
| Db | 5961 | GACCTGATTAATCTGTCTCCCTGCACTCTCTCTCTCTGCTGTGGCTGTGTGGGGTCTGTG | 6020 |
| QY | 1021 | TGCGGAGCGATATCTGCCGTGGGCACTGGGCTCAAGGGAGGGGGCTGTGCAGTGAATGAAC | 1080 |
| Db | 6021 | TGCCACACCAATATCTGCCCGCGGCACTGGGGCCCGGGAGAAAGGGGCTGTGCATGATGAAC | 6080 |
| QY | 1081 | CGGCTGATAGAGGTTGCGCTCGCGGGGGTAAACAATGTTTCCCCCAGCACAATATGTGCCAAG | 1140 |
| Db | 6081 | CGGCTGATAGAGGTTGCTTTCGCGGGGGTAAACAACGCTTCCCCCAGCACAATATGTGTGCTAG | 6140 |
| QY | 1141 | AGCAGCGCCGAGACAGTGTCACTAGATCCCTCTCCAGCTTTAATAATACCCAACTGTGTG | 1200 |
| Db | 6141 | AGCAGCGCCGAGAGCGCGCTGTGATCTAGATCCTCTTCAGCCTTTACATATCTAGCTGCTTG | 6200 |
| QY | 1201 | AAGAGGCTCCACCAAGTGAATTAACGAGGACTGTCCACGCGCCGTGCTCGGCTGTGACTA | 1260 |
| Db | 6201 | AAGAGGCTTCCACCAAGTGAATCAAGAAACATGTCTCAAGCATGTGCTCGGCTGTGACTA | 6260 |
| QY | 1261 | AGGAGTGTGGGACTGGAATATGCAAGATTTTGGCTGACTTAAGAACCTGTGCTCAAGTCC | 1320 |
| Db | 6261 | AGGAGTGTGGGACTGGAATATGCAAGTGTGCTGACTGACTTTTAAAGACCTGTGCTCAAGTCC | 6320 |
| QY | 1321 | AAGCTCCGCGGCAATTAACGGGAGTCCCTTTTCTCAATGCCAACGTGGGATACAAAGGG | 1380 |
| Db | 6321 | AAGATCATAGCCGCGGATTAACGGGAGTCCCTTTCTCTGTGTCAACGGGCTACAAAGGGA | 6380 |
| QY | 1381 | GTCGTGGCGGGAGACCGGCATCATGCGAGACCACTGTGATGTGGAGCACAATCAACCGGA | 1440 |
| Db | 6381 | GTCGTGGAGGGGGAGCGGCATCATGAGACCACTGTGCCATGTGGAGCACAATCAACCGGA | 6440 |
| QY | 1441 | CATGTCAAAAACGTTTCCATGAGGATGCTTGGGCTTAAAGACTGTGTAACATGTGCAT | 1500 |
| Db | 6441 | CATGTCAAAAGAACGTTTCCATGAGGATGCTTGGGCTTAAACCTGACGAAACAGTGCAT | 6500 |
| QY | 1501 | GGAACATTTCCCATCAACGCAATTAACCAACGGGCCCCCTGCAAGCCTTCCCAAGCGCCAAAC | 1560 |
| Db | 6501 | GGAACATTTCCCATCAACGCAATTAACCAACGGGCCCCCTGCAACCTTCCCGGCGCAAAAC | 6560 |
| QY | 1561 | TATTCACAGGGCGCTGTGGCGGGGTGCTGTAGAGATGCTGTGAGATTAAACGGGGTGGGG | 1620 |
| Db | 6561 | TATTCACAGGGCGCTGTGGCGGGGTGCTGTAGAGATGCTGTGAGATTAAACGGGGTGGGG | 6620 |
| QY | 1621 | GATTTCCACTACGTGACGAGCATGACCACTGACCAACGTTAAATGCGCCGTGCAGAGTTTCCA | 1680 |
| Db | 6621 | GATTTCCACTACGTGACGAGCATGACCACTGACCAACGTTAAATGCGCCATGCAAGTTTCCA | 6680 |
| QY | 1681 | GCCCCCGGAATCTTTCACAGAAGTGATGGGGTGGGCTGCAACAGGTACGCTTCGGCGTGC | 1740 |
| Db | 6681 | GCCCCCGGAATCTTTCACAGAAGTGATGGAGTACGAGTTGCAACAGGTACGCTTCGGCGTGC | 6740 |
| QY | 1741 | AAACCTCTCCACGGGAGAGAGGTCAANTTCAAGGTCCGGGCTCAACCAATACCTGGTTGGG | 1800 |
| Db | 6741 | AAACCTCTCTTACGGGAGAGAGGTCAANTTCAAGGTCCGGGCTCAACCAATATCTGGTTGGG | 6800 |
| QY | 1801 | TCGCAAGCTCCCATGCGAGCCGGAACCGGATGTAGCAGTGTCACTTCATGCTCAACCGAC | 1860 |
| Db | 6801 | TCACACATCTCCATGCGAGCCGGAACCGGATGTAGCAGTGTCACTTCATGCTCAACCGAC | 6860 |
| QY | 1861 | CCCTTCCACATCAACGACAGAGACGGCTTAAGCGCAGGCTGGCCAGGGGGTCTTCCCTCTCC | 1920 |
| Db | 6861 | CCATTCACATTAACGACAGAGACGGCTTAAGCGTAAAGGTAGGCTGGCCAGGGGGTCTTCCCTCTCC | 6920 |
| QY | 1921 | TTGGCCAGGCTCTTTCAGCTAGCCAGTGTGTGTGCGCTTCTGGAAGGCGACATTAATTAAC | 1980 |
| Db | 6921 | TTGGCCAGGCTCTTTCAGCTAGCCAGTGTGTGTGCGCTTCTTGAAGGCGACATTAATTAAC | 6980 |

| | | | |
|----------------------------|--|---|-----------------|
| QY | 1981 | CAAAATGACCTTCCAGAGCGCTGACCTATGAGGCCAACCTCGTGGCGCATGATG | 2045 |
| Db | 6981 | CATATGATCTCCCGAGCGCTGACCTATGAGGCCAACCTCTGTGGCGCAAGATG | 7040 |
| QY | 2041 | GGC 2043 | |
| Db | 7041 | GGC 7043 | |
| RESULT 12 | | | |
| AR005087 | | 7863 bp | DNA |
| LOCUS | | | linear |
| DEFINITION | Sequence 35 from patent US 5747339. | | PAT 04-DEC-1998 |
| ACCESSION | AR005087 | | |
| VERSION | AR005087.1 | GI:3965966 | |
| KEYWORDS | | | |
| SOURCE | Unknown. | | |
| ORGANISM | Unknown. | | |
| REFERENCE | 1 (bases 1 to 7863) | | |
| AUTHORS | Okayama,H., Fuke,I., Mori,C., Takamizawa,A. and Yoshida,I. | | |
| TITLE | Non-B hepatitis virus genomic cDNA and antigen polypeptide | | |
| JOURNAL | Patent: US 5747339-A 35 05-MAY-1998; | | |
| FEATURES | Location/Qualifiers | | |
| source | 1..7863 | | |
| ORIGIN | /organism="unknown" | | |
| | /mol_type="unassigned DNA" | | |
| Query Match | 87.9%; Score 1796.6; DB 6; Length 7863; | | |
| Best Local Similarity | 92.5%; Pred. No. 0; | | |
| Matches 1889; Conservative | 0; Mismatches 154; Indels 0; Gaps 0; | | |
| QY | 1 | TGGAGAGGCGCTTTTCAAGGCTTCAGCCACGCTGATGCCACTTCTGTCCCAACAAG | 60 |
| Db | 3505 | TGGAGAGAGTGTCTTCAAGGCGCTCACCCCATATAGATGACACTTCTGTCCAGACCAAG | 3564 |
| QY | 61 | CAGGACAGAGACACTTCCCCTACCTGTGGCGTACAGAGCTACTGTGTGCGCTAAGGACC | 120 |
| Db | 3665 | CAGGACAGAGACACTTCCCCTACTGTGTAGCATACAGACGAGTGTGCGCCAGGACT | 3624 |
| QY | 121 | CAGGCCCCACCTCCCATGAGGATCAATGTGAAAGTGTCTATACGCGCTAAAGCTTACT | 180 |
| Db | 3625 | CAGGCCCCACCTCCCATGAGGATCAATGTGAAAGTGTCTATACGCGCTAAAGCTTACG | 3684 |
| QY | 181 | CTGGCGGGGCGCAACCTCTGTGTATAGCTGGGAGCGCTGCAAAACAGAGTACCTCT | 240 |
| Db | 3685 | CTGCACGGGCGCAACCTCTGTGTATAGAGCTGGAGCCGCTGCGAATGAGTCACTCTC | 3744 |
| QY | 241 | ACACACCCCAATPACCAATTCATCATGAGGATGCAATGACCGACCTGGAGAGTGTCAAG | 300 |
| Db | 3745 | ACCCACCCCAATPACCAATTCATCATGAGGATGCAATGACCGCTGGAGAGTGTCACT | 3804 |
| QY | 301 | AGCAGCTGGGTGTGTGGGCGGGGCTCTTGACAGCTCTGCTGCGTATTGCTTGACACA | 360 |
| Db | 3805 | AGCAGCTGGGTGTGTGGGCGGGGCTCTTGACAGCTCTGCGCGGTATTGCTTGACACA | 3864 |
| QY | 361 | GGCAGCGTGTCTATTTGGGTAGGATATCTTGTCCGGGCGGCGCGTATTGTTCCTCGAC | 420 |
| Db | 3865 | GGCAGTGTGTCTATTTGGGTAGGATATTCTTGTCCGGGAGGCGCGGCAATTGTTCCTCGAC | 3924 |
| QY | 421 | AGGAGAGTCTCTACAGAGATTGATGAGATGGAAAGTGGCGGTGCGACCTCCCTTAC | 480 |
| Db | 3925 | AGGAGAGTCTCTACAGAGATTGATGAGATGGAAAGTGGCGGTGCGACCTCCCTTAC | 3984 |
| QY | 481 | ATCGAGCAGGGAATGACGCTTCGCCAGACAGTTCAAGCAAAAAGCGCTCGAGTGTCTGAC | 540 |
| Db | 3985 | ATCGAGCAGGGAATGACGCTTCGCCAGACAAATTCAGCAGAAAAGCGCTCGAGTTACTCAA | 4044 |
| QY | 541 | ACAGCCACCAAGCAGAGGAGCGCGTGTCCGTTGGATGAGTCCAAAGTGGCGAGCCCTT | 600 |
| Db | 4045 | ACAGCCACCAAGCAGAGGAGCGTGTCCGTTGGATGAGTCCAAAGTGGCGAGCCCTT | 4104 |

```

OY 601 GAGACCTTGTGGCGAACAATGTGAACTTCAATCAGCGGATACAGTACTTACAGGC 660
Db 4105 GAGACATTTCTGGCGAAGACATGTGAAATTTTCATCAGCGGATACAGTACTTACAGGC 4164
OY 661 TTGTCCACTGTGCGGGGAATCCGGGATGTGCATCAGTATGGCGCTTCAACGCTCTGTC 720
Db 4165 TTATCACTTGTCTGGGAACCCCGCAATAGCATATGATGATTCACAGCTCTATC 4224
OY 721 ACTAGCCGCTCACCACCCCAATCTACCTCTCTTAAATCTGCGGGGATGGTAAAGC 780
Db 4225 ACGACCCGCTCACCACCCCAAGTACCTCTCTTAAATCTTGGGGGGTGGGTGGCT 4284
OY 781 GCCCACTGTCTCCCCCACTGTCTGCTTCTGTAAGGCGCGCATCTCTGTGTCG 840
Db 4285 GCCCACTGTCTCCCCCACTGTCTGCTTCTGTAAGGCGCGCATCTCTGTGTCG 4344
OY 841 GCTGTGGAGAGATAGGCTTGGGAAGGTCTGTGGAATCTTGGCGGGATGAGAGA 900
Db 4345 GCTGTGGAGAGATAGGCTTGGGAAGGTCTGTGGAATCTTGGCGGGATGAGAGA 4404
OY 901 GGAGTGGAGGCGCGCTGTGAGCTTTTAAAGTATAGAGCGGAAATGCCCTCCACGAG 960
Db 4405 GGAGTGGCGCGCGCTGTGAGCTTTTAAAGTATAGAGCGGAGATGCCCTCCACGAG 4464
OY 961 GACCTGTCTTACTTACTCTCTGCTGCTCTCTCTCTGCTGTGCTGTGCTGTGCT 1020
Db 4465 GACCTGTCTTACTTACTCTCTGCTGCTCTCTCTCTGCTGTGCTGTGCTGTGCT 4524
OY 1021 TGGCGAGGATCTAGTGTGAGCGAGTGGGTCCAGGGGAGGGGGTGTGAGTGAAG 1080
Db 4525 TGTGAGCAATATCTGCTGTGAGCACTGTGGTCCGGAGAGGGGGCTGTGAGTGAAG 4584
OY 1081 CGGCTGATAGCTTGTGCTGTGAGGAGTAACTATGTTTCCCGCACGCACTATGTGCAAG 1140
Db 4585 CGGCTGATAGCTTGTGCTGTGAGGAGTAACTATGTTTCCCGCACGCACTATGTGCTAG 4644
OY 1141 AGCGAGCGCGACGACGTGTCTACTAGATCTTCTCCAGCTTATCTATCACTCACTGTTG 1200
Db 4645 AGCGAGCGCGACGCGCTGTCTACTAGATCTTCTCCAGCTTATCTATCACTCACTGCTG 4704
OY 1201 AAGAGGCTCCACCAATGATTAACAGAGATGTCTCCAGGCGCTGTGCTGTGCTGTA 1260
Db 4705 AAGAGGCTCCACCAATGATTAACAGAGATGTCTCCAGGCGCTGTGCTGTGCTGTA 4764
OY 1261 AGGAGTGTGGAGCTGATATGACAAGTGTGTGCTGACTTCAAGACTGTGCTCAAGTCC 1320
Db 4765 AGGAGTGTGGAGCTGATATGACAAGTGTGTGCTGACTTCAAGACTGTGCTCAAGTCC 4824
OY 1321 AAGCTCTGTGCGGATTAACCGGAGTCCCTTTTCTCATGTCCAACTGTGGTACAAAGGG 1380
Db 4825 AAGCTCTGTGCGGATTAACCGGAGTCCCTTTTCTCATGTCCAACTGTGGTACAAAGGG 4884
OY 1381 GTCTGCGGGGAGAGCGGATTCATGAGACACCTGTCTATGTGAGCAAGATCACCGGA 1440
Db 4885 GTCTGCGGGGAGAGCGGATTCATGAGACACCTGTCTATGTGAGCAAGATCACCGGA 4944
OY 1441 CATGTCAAAAACGCTTCCATGAGATCGTTGGGCTTAAGACTGTATGTAACATGTGCAAT 1500
Db 4945 CATGTCAAAAACGCTTCCATGAGATCGTTGGGCTTAAGACTGTATGTAACATGTGCAAT 5004
OY 1501 GGAATCTTCCCATCAACGATACACACGCGGCCCTGTGACGCGCTCCAGCGCCAAAC 1560
Db 5005 GGAATCTTCCCATCAACGATACACACGCGGCCCTGTGACGCGCTCCAGCGCCAAAC 5064
OY 1561 TATTCAGGGGCTGTGAGGGGTGTGCTGTGAGAGTATGTAAGGTTTAAGGGGGTGGG 1620
Db 5065 TATTCAGGGGCTGTGAGGGGTGTGCTGTGAGAGTATGTAAGGTTTAAGGGGGTGGG 5124
OY 1621 GATTTCACTAGTACGAGCATGACCACTGCAACAGTAAATGCCCCTGCGCAGGTTCCA 1680
Db 5125 GATTTCACTAGTACGAGCATGACCACTGCAACAGTAAATGCCCCTGCGCAGGTTCCG 5184
OY 1681 GCCCCGAATCTTCAAGAGTGTGAGGGGTGCGGCTGCACAGTACGCTCCGCGCTGC 1740

```

```

Db 5185 GCTCTGATTTCTTCCGAGGTGAGCGAGTGCCTTCCACAGTACGCTTCCGCGTGC 5244
OY 1741 AAACCTCTCTTACGGGAGAGGTCAATTCAGTGTGGGCTCAACCAATATCTGTGG 1800
Db 5245 AGGCTCTCTTACGGGAGAGGTCAATTCAGTGTGGGCTCAACCAATATCTGTGG 5304
OY 1801 TCGAGCTTCCATGCGAGCCCGAACCAGATGTAGAGTGTCTCATCTTCCATGTCACGAC 1860
Db 5305 TCAGAGCTTCCATGCGAGCCCGAACCAGATGTAGAGTGTCTCATCTTCCATGTCACGAC 5364
OY 1861 CCCTCCCATCAACGAGAGCGGCTAAGGCGAGGCTGTGCGAGGGGTCTCCCGCTCC 1920
Db 5365 CCCTCCCATCAACGAGAGCGGCTAAGGCGAGGCTGTGCGAGGGGTCTCCCGCTCC 5424
OY 1921 TTGGCAGCTCTTACGATGACGATGTGTGCTGCTGCTGCTGAGAGCGACATATATAC 1980
Db 5425 TTGGCAGCTCTTACGATGACGATGTGTGCTGCTGCTGCTGAGAGCGACATATAC 5484
OY 1981 CAATATGACTTCCAGAGCGCTGACTCATTCGAGGCCAACCTCTGTGGCGCATGAGATG 2040
Db 5485 CACCATGTCTCTCCGAGCGTGAACCTCATTCAGAGGCCAACCTCTGTGGCGCATGAGATG 5544
OY 2041 GGC 2043
Db 5545 GGC 5547

RESULT 13
AR064526 7863 bp DNA linear PAT 29-SEP-1999
LOCUS AR064526 Sequence 35 from patent US 5847101.
DEFINITION AR064526
ACCESSION AR064526
VERSION AR064526.1 GI:5993834
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 7863)
AUTHORS Okayama,H., Fuke,I., Mori,C., Takamizawa,A. and Yoshida,I.
TITLE Non-A, non-B hepatitis virus genomic cDNA and antigen polypeptide
JOURNAL Patent: US 5847101-A 35 08-DEC-1998;
FEATURES
location/Qualifiers
source
1..7863
/organism="Unknown"
/mol_type="unassigned DNA"

ORIGIN
Query Match 87.9%; Score 1796.6; DB 6; Length 7863;
Best Local Similarity 92.5%; Pred. No. 0;
Matches 1889; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

OY 1 TGGGAGGGGCGTCTTACAGAGGCTCTCACCGCATGATGCCACTTCTGTCCAAACAAG 60
Db 3505 TGGGAGGATGTTTACAGGCTCTCACCATATGATGACACTTCTGTCCAGACCAAG 3564
OY 61 CAGGACGAGACCAATTCCTCCCTACCTGTGTGGGTAACAAGTACTGTGTGCGTTAGGGCC 120
Db 3565 CAGGACGAGACCAATTCCTCCCTACCTGTGTGGGTAACAAGCAACGATGTGCGCGAGGCT 3624
OY 121 CAGGCCCCACCTCCATCATGAGATCAAAATGTGGAAGTGTCTCATCGCTTAAGCTACT 180
Db 3625 CAGGCCCCACCTCCATCATGAGATCAAAATGTGGAAGTGTCTCATCGCTTAAGCTACT 3684
OY 181 CTGGCGGGGCAACAACCTTGTGTATATGAGGTGGAGCGGTCCAAAGAGAGTCAACCTC 240
Db 3685 CTGGCGGGGCAACAACCTTGTGTATATGAGGTGGAGCGGTCCAAAGAGAGTCAACCTC 3744
OY 241 ACACACCCCATTAACCAATTCATCATGTGATGATGATGATGATGATGATGATGATGATG 300
Db 3745 ACCACCCCATTAACCAATTCATCATGTGATGATGATGATGATGATGATGATGATGATG 3804
OY 301 AGCACCTGGGTGCTGTGTGGGCGGGGCTCTTGACGCTGTGCTGTGATTTGCTTGACACA 360

```


Db 3505 TGGGAGAGTGTCTTACAGAGGCTCACCCATATAGACACTTCTGTCCAGACCAAG 3564
Qy 61 CAGGAGGAGAGCAACTTCCCTTACCTGTGTGGGATACAGGCTATCTGTGTGCGCTAGGCG 120
Db 3565 CAGGAGGAGAGCAACTTCCCTTACCTGTGTGGGATACAGGCTATCTGTGTGCGCTAGGCG 3624
Qy 121 CAGGAGGAGAGCAACTTCCCTTACCTGTGTGGGATACAGGCTATCTGTGTGCGCTAGGCG 180
Db 3625 CAGGAGGAGAGCAACTTCCCTTACCTGTGTGGGATACAGGCTATCTGTGTGCGCTAGGCG 3684
Qy 181 CTGGCGGGGCGCAACCCCTTGTGTATAGGCTGGGAGCGGCTCAAAACGAGGCTCACCCCTC 240
Db 3685 CTGGCGGGGCGCAACCCCTTGTGTATAGGCTGGGAGCGGCTCAAAACGAGGCTCACCCCTC 3744
Qy 241 ACACACCCCATTAACAAATTCATCATGTGCATGTCAAGCCGACTGGAGGTCGTCAAG 300
Db 3745 ACCGACCCCATTAACAAATTCATCATGTGCATGTCAAGCCGACTGGAGGTCGTCAAG 3804
Qy 301 AGCAGCTGGGTGTGTGTGGGCGGGGTCTTGTGAGCTGTGGCTGCTGATTTGCTTGAACA 360
Db 3805 AGCAGCTGGGTGTGTGTGGGCGGGGTCTTGTGAGCTGTGGCTGCTGATTTGCTGACA 3864
Qy 361 GGCAGCGGTGTGTGTGTGGGTAGATCATCTGTCCGGGCGGCGGCTATTTGTCCGAGC 420
Db 3865 GGCAGGTGTGTGTGTGGGTAGATCATCTGTCCGGGCGGCGGCTATTTGTCCGAGC 3924
Qy 421 AGGGAAGTCTCTTACAGAGAGTTGATGATGAGAGAGTGTGCGCTGCACTCCCTTAC 480
Db 3925 AGGGAAGTCTCTTACAGAGAGTTGATGATGAGAGAGTGTGCGCTGCACTCCCTTAC 3984
Qy 481 ATCGAGCAGGAGATGAGTGTGCGGCAAGCTTCAAGCAAAAGCGCTGGGTGTGCTGAG 540
Db 3985 ATCGAGCAGGAGATGAGTGTGCGGCAAGCTTCAAGCAAAAGCGCTGGGTGTGCTGAG 4044
Qy 541 ACAGCAGCAGAGAGAGGCGGCTGTCCGGGTGTGGAGTCCAAAGTGGAGAGCCCT 600
Db 4045 ACAGCAGCAGAGAGAGGCGGCTGTCCGGGTGTGGAGTCCAAAGTGGAGAGCCCT 4104
Qy 601 GAGACCTTCTGGGCGAAACACATGTGAATTCATCAGCGGATACAGTACTTACAGAGC 660
Db 4105 GAGACCTTCTGGGCGAAACACATGTGAATTCATCAGCGGATACAGTACTTACAGAGC 4164
Qy 661 TTGTCACTCTGCTGTGGGATCCCGGATGTGATCATCTGATGCGCTTCAAGCTTCTGTC 720
Db 4165 TTGTCACTCTGCTGTGGGATCCCGGATGTGATCATCTGATGCGCTTCAAGCTTCTGTC 4224
Qy 721 ACTAGCGGCTACACCAATCTAATCCCTCGCTTAACTCTGGGGGATGGGATGCG 780
Db 4225 ACTAGCGGCTACACCAATCTAATCCCTCGCTTAACTCTGGGGGATGGGATGCG 4284
Qy 781 GCCCACTGTGCTCCCGAGTGTGCTTGTGATGAGCGCGGCAATTTCTGTGTGCG 840
Db 4285 GCCCACTGTGCTCCCGAGTGTGCTTGTGATGAGCGCGGCAATTTCTGTGTGCG 4344
Qy 841 GCTGTGTGAGCATAGAGCTTGGAAAGTGTGCTTGTGATCTTGTGGCGGCTATGAGAGA 900
Db 4345 GCTGTGTGAGCATAGAGCTTGGAAAGTGTGCTTGTGATCTTGTGGCGGCTATGAGAGA 4404
Qy 901 GGAAGTGGAGAGGCGGCTGTGGCTTAAAGTCAAGAGGGGAAATGCCCTCCACCGAG 960
Db 4405 GGAAGTGGAGAGGCGGCTGTGGCTTAAAGTCAAGAGGGGAAATGCCCTCCACCGAG 4464
Qy 961 GACCTGTAACTTACTCTGCTGCACTCTCTCTGCTGTGCTGGTCTGCGGAGCTGTG 1020
Db 4465 GACCTGTAACTTACTCTGCTGCACTCTCTCTGCTGTGCTGGTCTGCGGAGCTGTG 4524
Qy 1021 TGGCAGCGATATCTGCTGCGCATCTGTGCTCAAGGGAGAGGAGCTGTGCAATGTGAAC 1080
Db 4525 TGGCAGCGATATCTGCTGCGCATCTGTGCTCAAGGGAGAGGAGCTGTGCAATGTGAAC 4584
Qy 1081 CCGGCTGATGCGCTTGGCTTGGGGGATACATGTTTCCCGACGCACTATGTGCCAGAG 1140
Db 4585 CCGGCTGATGCGCTTGGCTTGGGGGATACATGTTTCCCGACGCACTATGTGCCAGAG 4644

Qy 1141 AGCGAGCGCGAGCAGCTGTCACTGATTCCTCTTCCGACTTTATATACCCAACTGTG 1200
Db 4645 AGCGAGCGCGAGCAGCTGTGTACTGATTCCTCTTCCGACTTTATATACCCAACTGTGCTG 4704
Qy 1201 AAGAGGCTCCACAGTGGATTTAAGAGAGCTGTCCAGCCCTGTCCGGCTCGTGTGCTA 1260
Db 4705 AAGAGGCTCCACAGTGGATTTAAGAGAGCTGTCCAGCCCTGTCCGGCTCGTGTGCTA 4764
Qy 1261 AAGATGTTTGGAGCTGATATGACAGTGTGTGCTGACTTCAAGACTGTGCTCAAGTCC 1320
Db 4765 AAGATGTTTGGAGCTGATATGACAGTGTGTGCTGACTTCAAGACTGTGCTCAAGTCC 4824
Qy 1321 AAGCTCTGCGCGATTAACGGAGAGTCCCTTTTCTATGCTCAAGCTGTGCTCAAGAGG 1380
Db 4825 AAGCTCTGCGCGAGCTACCTGAGTCCCTTTTCTGCTGCTCAAGCGGGGTACAGAGGA 4884
Qy 1381 GTCTGGGGGGAGAGCGGATCATGAGAGCAACCTGCTATGTGAGCAGATCACCGGA 1440
Db 4885 GTCTGGGGGGAGAGCGGATCATGAGAGCAACCTGCTATGTGAGCAGATCACCGGA 4944
Qy 1441 CATGTCAAAAACGCTTCATGAGATCGTTGGGCTTAAGACTGTATGATGATGAGCAT 1500
Db 4945 CATGTCAAAAACGCTTCATGAGATCGTTGGGCTTAAGACTGTATGATGATGAGCAT 5004
Qy 1501 GGAACATTCCTCCATCAACGATACACAGCGGCTTGCAGGCTTCCCGAGCGCAAC 1560
Db 5005 GGAACATTCCTCCATCAACGATACACAGCGGCTTGCAGGCTTCCCGAGCGCAAC 5064
Qy 1561 TATTCAGAGGGGCTGTGGCGGGGTGTGCTGAGAGTACGTGAGGTAAAGCGGGGTGGG 1620
Db 5065 TATTCAGAGGGGCTGTGGCGGGGTGTGCTGAGAGTACGTGAGGTAAAGCGGGGTGGG 5124
Qy 1621 GATTTCACTACGTGACAGAGATGACACTGACAACTGTAATGCTGCTGCGAGTTTCA 1680
Db 5125 GATTTCACTACGTGACAGAGATGACACTGACAACTGTAATGCTGCTGCGAGTTTCCG 5184
Qy 1681 GCCCGGATTTCTTCAAGAAAGTGTGAGGTGTGCGGCTGTGACAGGTATGCTTCCGGCTGTC 1740
Db 5185 GCTCTGAATTTCTTCTCGAGAGTGTGACGAGGTGTGACAGGTATGCTTCCGGCTGTC 5244
Qy 1741 AAACCTCTCTAAGGAGAGAGGTGACATTTCAAGGTGCGGCTTCAACCAATCTGCTGTGG 1800
Db 5245 AAGCTCTCTCTAAGGAGAGAGGTGACATTTCAAGGTGCGGCTTCAACCAATCTGCTGTGG 5304
Qy 1801 TCGAGCTCCCATGCGAGCCCGAACCAGATGAGCAGTGTCACTTCCATGCTCACCGAC 1860
Db 5305 TCGAGCTCCCATGCGAGCCCGAACCAGATGAGCAGTGTCACTTCCATGCTCACCGAC 5364
Qy 1861 CCTTCCCATCAACGAGAGAGCGGCTTAAAGCGAGGCTGTGCCAGGGGTCTTCCCGCTTCC 1920
Db 5365 CCTTCCCATCAACGAGAGAGCGGCTTAAAGCGAGGCTGTGCCAGGGGTCTTCCCGCTTCC 5424
Qy 1921 TTGGCAGCTCTTTCAGCTAGCCAGTGTGTGAGCTTCTGTAAGGAGCATATCACTTACC 1980
Db 5425 TTGGCAGCTCTTTCAGCTAGCCAGTGTGTGAGCTTCTGTAAGGAGCATATCACTTACC 5484
Qy 1981 CAATATGACTTCCAGAGAGCTGACCTCATGAGGCAACTCTGTGTGCGGAGTGAATG 2040
Db 5485 CACCATGTCTCTCCGAGAGCTGACCTCATGAGGCAACTCTGTGTGCGGAGAGTGAATG 5544
Qy 2041 GGC 2043
Db 5545 GGC 5547

RESULT 15
ARI45568
LOCUS ARI45568 7863 bp DNA linear PAT 08-AUG-2001
DEFINITION Sequence 35 from patent US 6217872.
ACCESSION ARI45568
VERSION ARI45568.1 GI:15108757
KEYWORDS

SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 7863)
AUTHORS Okayama,H., Fuke,I., Mori,C., Takamizawa,A. and Yoshida,I.
TITLE Non-A, non-B hepatitis virus genomic cDNA and antigen polypeptide
JOURNAL Patent: US 6217872-A 35 17-APR-2001;
FEATURES Location/Qualifiers
source 1..7863
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN

Query Match 87.9%; Score 1796.6; DB 6; Length 7863;
Best Local Similarity 92.5%; Pred. No. 0;
Matches 1889; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

QY 1 TGGAGGGCGCTCTTCAAGGCTTCAAGCCAGCTGATGCCACTTCTGTCCGAAACAAAG 60
DB 3505 TGGAGAGTGTCTTCAAGGCTTCAAGCCAGCTGATGCCACTTCTGTCCGAAACAAAG 3564

QY 61 CAGGAGAGAGCAACTTCCCTTAAGCTGTGGCGTACAGGCTAAGTGTGTGGCGTGGGCC 120
DB 3565 CAGGAGAGAGCAACTTCCCTTAAGCTGTGGCGTACAGGCTAAGTGTGTGGCGTGGGCC 3624

QY 121 CAGGCGCCCACTTCATGAGGATCAAAATGTGAAAGTGTCTCATACGGCTAAAGCTTACT 180
DB 3625 CAGGCGCCCACTTCATGAGGATCAAAATGTGAAAGTGTCTCATACGGCTAAAGCTTACT 3684

QY 181 CTGCGCGGGCCAAACCTTGTGTATAGAGCTGGAGCGGTCCAAACAGAGTCAACCTTC 240
DB 3685 CTGCGCGGGCCAAACCTTGTGTATAGAGCTGGAGCGGTCCAAACAGAGTCAACCTTC 3744

QY 241 ACAACACCCCAATACCAATATTCATGAGCATGATGTCAAGCGACTGAGAGTGTCTACG 300
DB 3745 ACCCAACCCCAATACCAATATTCATGAGCATGATGTCTGAGCTGAGAGTGTCTACG 3804

QY 301 AGCACTGGGTCGTGTGGCGGGGCTTTCAGAGCTGCTGCGATGTCTTTCAGCAAA 360
DB 3805 AGCACTGGGTCGTGTGGCGGGGCTTTCAGAGCTGCTGCGATGTCTTTCAGCAAA 3864

QY 361 GGCAGCGTGTGATTTGTGGGTAGATCATCTTGTCCGGCGGCGCTATTGTTCCGAC 420
DB 3865 GGCAGCGTGTGATTTGTGGGTAGATCATCTTGTCCGGCGGCGCTATTGTTCCGAC 3924

QY 421 AGGGAAGTCTCTTACCAAGAGTTCCATGAGATGAAAGAGTGGCGTCCGACCTCCCTTAC 480
DB 3925 AGGGAAGTCTCTTACCAAGAGTTCCATGAGATGAAAGAGTGGCGTCCGACCTCCCTTAC 3984

QY 481 ATTCGAGAGGGAATGCAAGTCCGAGCAAGTTCAGCAAAAGCGCTCGGTTGCTGCAG 540
DB 3985 ATTCGAGAGGGAATGCAAGTCCGAGCAAGTTCAGCAAAAGCGCTCGGTTGCTGCAG 4044

QY 541 ACAGCCACCAAGCAAGCGAGGCGCTGCTCCGTTGGTGAAGTCCAAAGTGGCGAGCCCTT 600
DB 4045 ACAGCCACCAAGCAAGCGAGGCGCTGCTCCGTTGGTGAAGTCCAAAGTGGCGAGCCCTT 4104

QY 601 GAGACCTTCTGGGCGAAACATGTRGAATTCATCAGCGGGAATCACTTACAGAGC 660
DB 4105 GAGACCTTCTGGGCGAAACATGTRGAATTCATCAGCGGGAATCACTTACAGAGC 4164

QY 661 TTGTCACTCTGGCGGGAATCCCGGATTTGATCATGAGCTTTCACACCTCTGTC 720
DB 4165 TTGTCACTCTGGCGGGAATCCCGGATTTGATCATGAGCTTTCACACCTCTGTC 4224

QY 721 ACTAGCCGCTACACCAATCTACCTCTGCTTAAATCATCTGGGGGATGAGTACG 780
DB 4225 ACTAGCCGCTACACCAATCTACCTCTGCTTAAATCATCTGGGGGATGAGTACG 4284

QY 781 GCCCACTGTGCTCCCGCAGTGTGCTTCACTTTCTGTAAGCGCGCGGCAATGCTGTGCG 840
DB 4285 GCCCACTGTGCTCCCGCAGTGTGCTTCACTTTCTGTAAGCGCGCGGCAATGCTGTGCG 4344

QY 841 GCTGTGGCAGATAGGCTTGGGAAGGTGTTGTGACATCTTGGCGGGATATGAGCA 900
DB 4345 GCTGTGGCAGATAGGCTTGGGAAGGTGTTGTGACATCTTGGCGGGATATGAGCA 4404

QY 901 GGATGTGGCAGGCGCGCTCGTGCCTTTTAAAGTATAGAGCGGAAATGCCCTTCACCGAG 960
DB 4405 GGATGTGGCAGGCGCGCTCGTGCCTTTTAAAGTATAGAGCGGAAATGCCCTTCACCGAG 4464

QY 961 GACCTGTAACTTACTCCCTGCAATCCCTCTCTCCGCTGAGCTGTGCTGCGGGGCTG 1020
DB 4465 GACCTGTAACTTACTCCCTGCAATCCCTCTCTCCGCTGAGCTGTGCTGCGGGGCTG 4524

QY 1021 TGGCAGCAGATACCTCGTGGGACAGTGGGTCCAGGGAGAGGGGCTGTGAGTGAATC 1080
DB 4525 TGGCAGCAGATACCTCGTGGGACAGTGGGTCCAGGGAGAGGGGCTGTGAGTGAATC 4584

QY 1081 CGGCTGATAGCGTTGCGCTCGCGGGGTAAACATGTTTCCCGACGCACTATGTGCCAG 1140
DB 4585 CGGCTGATAGCGTTGCGCTCGCGGGGTAAACATGTTTCCCGACGCACTATGTGCCAG 4644

QY 1141 AGCGAGCGGAGCAGTGTCTCATGATTCCTCTCCGACCTTAATACCCAACTGTG 1200
DB 4645 AGCGAGCGGAGCAGTGTCTCATGATTCCTCTCCGACCTTAATACCCAACTGTG 4704

QY 1201 AAGAGGCTCCACAGTGAATTAACGAGAGCTGCTCCAGCGCTGCTCCGGCTGTAGCTA 1260
DB 4705 AAGAGGCTCCACAGTGAATTAACGAGAGCTGCTCCAGCGCTGCTCCGGCTGTAGCTA 4764

QY 1261 AGGAGTGTGGAGCTGATATGACATGTTTGGCTGACTTCAAGCCTGAGCTCCAGTCC 1320
DB 4765 AGGAGTGTGGAGCTGATATGACATGTTTGGCTGACTTCAAGCCTGAGCTCCAGTCC 4824

QY 1321 AAGCTCTGCGCGCAATACCGGAGTCCCTTTTCTCATGCGCAAGTGGATCAAGGGG 1380
DB 4825 AAGCTCTGCGCGCAATACCGGAGTCCCTTTTCTCATGCGCAAGTGGATCAAGGGG 4884

QY 1381 GTCTGGCGGGAGAGCGGATCATGACAGCACTGTCTCATGTGAGCAGATACCGGA 1440
DB 4885 GTCTGGCGGGAGAGCGGATCATGACAGCACTGTCTCATGTGAGCAGATACCGGA 4944

QY 1441 CATGTCAAAAACGGTTCCATGAGATCGTTGGGCTTAAGACCTGTATGAATGTGGCAT 1500
DB 4945 CATGTCAAAAACGGTTCCATGAGATCGTTGGGCTTAAGACCTGTATGAATGTGGCAT 5004

QY 1501 GGAATTTCCCATTAACGCAATACCAACAGGGGCTTGCAGCCCTCCCAACGCCCAAC 1560
DB 5005 GGAATTTCCCATTAACGCAATACCAACAGGGGCTTGCAGCCCTCCCAACGCCCAAC 5064

QY 1561 TATTCAGGGCGCTGTGGCGGGTGGCTGTGAGAGTACGTGAAGGTTACGGGGTGGG 1620
DB 5065 TATTCAGGGCGCTGTGGCGGGTGGCTGTGAGAGTACGTGAAGGTTACGGGGTGGG 5124

QY 1621 GATTTCCACTACGTGACGAGCATGACCACTGCAAAAGTAAATGCGCGGTGCCAGTTCCA 1680
DB 5125 GATTTCCACTACGTGACGAGCATGACCACTGCAAAAGTAAATGCGCGGTGCCAGTTCCA 5184

QY 1681 GCCCGCGAATTTCTTCAAGAAAGTGAAGTGGGCTGCGCTCAAGAGTACCTCCGGGTG 1740
DB 5185 GCCCGCGAATTTCTTCAAGAAAGTGAAGTGGGCTGCGCTCAAGAGTACCTCCGGGTG 5244

QY 1741 AAACCTCTCCTACGGGAGAGGTCATTCAGAGTCCGGGCTCAACCAATACCTGGTTGG 1800
DB 5245 AAACCTCTCCTACGGGAGAGGTCATTCAGAGTCCGGGCTCAACCAATACCTGGTTGG 5304

QY 1801 TGCAGCTTCCATGCGAGGCCGAAACCGGATGTAGAGTGTCTCATCTTCATGCTCACGAC 1860
DB 5305 TGCAGCTTCCATGCGAGGCCGAAACCGGATGTAGAGTGTCTCATCTTCATGCTCACGAC 5364

QY 1861 CCTTCCCATATACAGCAGAGCGGCTAAGCCAGGCTGAGCCAGGGGGTCTCCCTCTCC 1920
DB 5365 CCTTCCCATATACAGCAGAGCGGCTAAGCCAGGCTGAGCCAGGGGGTCTCCCTCTCC 5424

QY 1921 TTGGCAGGCTTTCAGCTAGCCAGTTGTGTGCGCTTCTCTGAAGGCGACATATAC 1980

| | | | |
|----|------|---|------|
| Db | 5425 | TTGGCCAGCTCTTCAGCTAGCCAGTTGTCGCGCTTCCTTGAGGCGCATGCACTACC | 5484 |
| Qy | 1981 | CAAAATGACTTCCAGACGCTGACCTCATCGAGGCCAACCTCCTGTGGCGCATGAGATG | 2040 |
| Db | 5485 | CACCATGTCTCTCCGAGCGCTGACCTCATCGAGGCCAACCTCCTGTGGCGCATGAGATG | 5544 |
| Qy | 2041 | GGC 2043 | |
| Db | 5545 | GGC 5547 | |

Search completed: February 25, 2005, 05:54:54
 Job time : 8561 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2005, 01:22:04 ; Search time 957 Seconds
(without alignments)
12637.442 Million cell updates/sec

Title: US-09-664-363-20

Perfect score: 2043

Sequence: 1 TGGGAGGCGCTTCACAGC.....TGTGGCGCATGAGATGGGC 2043

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_16Dec04:*

1: geneseqn19808:*

2: geneseqn19908:*

3: geneseqn20008:*

4: geneseqn20018:*

5: geneseqn20018:*

6: geneseqn20028:*

7: geneseqn20028:*

8: geneseqn20038:*

9: geneseqn20038:*

10: geneseqn20038:*

11: geneseqn20038:*

12: geneseqn20048:*

13: geneseqn20048:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|----------|---------------------|
| 1 | 2041.4 | 99.9 | 3750 | AAQ12241 | AaQ12241 Encodes P |
| 2 | 2033.4 | 99.5 | 7065 | AAQ61195 | AaQ61195 PT-NANBH |
| 3 | 1803 | 88.3 | 5955 | ACA61114 | ACA61114 HCV CDNA |
| 4 | 1803 | 88.3 | 5965 | ACA61111 | ACA61111 HCV CDNA |
| 5 | 1803 | 88.3 | 37090 | ACA61113 | ACA61113 Adenovirus |
| 6 | 1801.4 | 88.2 | 7987 | ADR38452 | Ad38452 DNA encod |
| 7 | 1801.4 | 88.2 | 9587 | ADR82189 | Ad82189 Hepatitis |
| 8 | 1796.6 | 87.9 | 9416 | AAQ20268 | AaQ20268 Non-A, no |
| 9 | 1796.6 | 87.9 | 9416 | AAQ21829 | AaQ21829 Non-A, no |
| 10 | 1796.6 | 87.9 | 9416 | AAQ59394 | AaQ59394 Non-A, no |
| 11 | 1787 | 87.5 | 7987 | AAQ25321 | AaQ25321 Hepatitis |
| 12 | 1787 | 87.5 | 7989 | AAQ98968 | AaQ98968 Hepatitis |
| 13 | 1787 | 87.5 | 7989 | AAQ25322 | AaQ25322 Hepatitis |
| 14 | 1787 | 87.5 | 7989 | AAQ57845 | AaQ57845 HCV repl1 |
| 15 | 1787 | 87.5 | 7992 | AAQ47276 | AaQ47276 Hepatitis |
| 16 | 1787 | 87.5 | 8001 | AAQ98967 | AaQ98967 Hepatitis |
| 17 | 1787 | 87.5 | 8637 | AAQ98966 | AaQ98966 Hepatitis |
| 18 | 1787 | 87.5 | 8638 | ABK88595 | ABK88595 Hepatitis |
| 19 | 1787 | 87.5 | 8639 | ABK88573 | ABK88573 Hepatitis |
| 20 | 1787 | 87.5 | 8649 | AAQ98969 | AaQ98969 Hepatitis |

| | | | | | | |
|----|--------|------|-------|----|----------|--------------------|
| 21 | 1787 | 87.5 | 9605 | 6 | ABK91424 | ABK91424 Hepatitis |
| 22 | 1787 | 87.5 | 9605 | 6 | ABK91411 | ABK91411 Hepatitis |
| 23 | 1787 | 87.5 | 9605 | 6 | ABK91425 | ABK91425 Hepatitis |
| 24 | 1787 | 87.5 | 9605 | 6 | ABK91426 | ABK91426 Hepatitis |
| 25 | 1787 | 87.5 | 10690 | 6 | ABK91422 | ABK91422 Hepatitis |
| 26 | 1787 | 87.5 | 10690 | 6 | ABK91412 | ABK91412 Hepatitis |
| 27 | 1787 | 87.5 | 10690 | 9 | ACA61697 | ACA61697 Hepatitis |
| 28 | 1787 | 87.5 | 10690 | 10 | ADC83762 | ADC83762 PHCVNeol7 |
| 29 | 1787 | 87.5 | 10691 | 6 | ABK91423 | ABK91423 Hepatitis |
| 30 | 1787 | 87.5 | 11076 | 3 | AAQ98965 | AaQ98965 Hepatitis |
| 31 | 1787 | 87.5 | 11313 | 12 | ADP86264 | ADP86264 Hepatitis |
| 32 | 1787 | 87.5 | 11313 | 12 | ADP86271 | ADP86271 Hepatitis |
| 33 | 1786.6 | 87.4 | 8642 | 6 | ABK88574 | ABK88574 Hepatitis |
| 34 | 1785.4 | 87.4 | 7141 | 6 | AAQ25333 | AaQ25333 Hepatitis |
| 35 | 1785.4 | 87.4 | 7789 | 6 | AAQ25330 | AaQ25330 Hepatitis |
| 36 | 1785.4 | 87.4 | 7980 | 10 | ADD93724 | ADD93724 Hepatitis |
| 37 | 1785.4 | 87.4 | 7980 | 10 | ADD93725 | ADD93725 Hepatitis |
| 38 | 1785.4 | 87.4 | 7987 | 6 | AAQ25329 | AaQ25329 Hepatitis |
| 39 | 1785.4 | 87.4 | 7987 | 6 | AAQ25324 | AaQ25324 Hepatitis |
| 40 | 1785.4 | 87.4 | 7989 | 6 | AAQ25326 | AaQ25326 Hepatitis |
| 41 | 1785.4 | 87.4 | 7989 | 6 | AAQ25325 | AaQ25325 Hepatitis |
| 42 | 1785.4 | 87.4 | 7989 | 10 | ADD93726 | ADD93726 Hepatitis |
| 43 | 1785.4 | 87.4 | 7989 | 10 | ADD93733 | ADD93733 Hepatitis |
| 44 | 1785.4 | 87.4 | 7989 | 10 | ADD93734 | ADD93734 Hepatitis |
| 45 | 1785.4 | 87.4 | 7991 | 6 | AAQ47279 | AAQ47279 Hepatitis |

ALIGNMENTS

RESULT 1
AAQ12241
ID AAQ12241 standard; CDNA; 3750 BP.

XX AAQ12241;
AC
XX
DT 25-MAR-2003 (revised)
DT 17-SEP-1991 (first entry)
XX
DE Encodes portion of PT-NANBH viral non-structural protein.
XX
KW post-transfusional non-A, non-B hepatitis; virus; vaccine; ss.
XX
OS Non-A.
OS non-B hepatitis virus.
XX
PN GB2339245-A.
XX
PD 26-JUN-1991.
XX
PF 17-DEC-1990; 90GB-00027250.
XX
PR 18-DEC-1989; 89GB-00028562.
PR 27-FEB-1990; 90GB-00004414.
PR 03-MAR-1990; 90GB-00004814.
XX
PA (WELL) WELLCOME FOUND LTD.
PA (HIGH) HIGHFIELD P E.
XX
PI Highfield PE, Rodgers BC, Tedder RS, Barbara JAJ;
XX
DR WPI; 1991-187584/26.
XX
DR P-PSDB; AAR12599.
XX
PT Post-transfusional non-A non-B hepatitis poly:peptide(s) - and also DNA
XX and antibodies used in diagnostic assays and in vaccines.
XX
PS Claim 10; Page 88-97; 108pp; English.
XX
CC This sequence probably encodes viral non-structural proteins of the PT-
CC NANBH viral genome which are antigenic. It was isolated from serum of
CC humans infected by the virus. See also AAQ12236-40 and AAQ12242. (Updated
CC on 25-MAR-2003 to correct PA field.)

XX Sequence 3750 BP; 794 A; 1140 C; 1072 G; 744 T; 0 U; 0 Other;
SQ
Query Match 99.9%; Score 2041.4; DB 2; Length 3750;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2042; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGGAGGGGCTTTCACAGGCTCAACCACTGATGCCACTTCTGTCCCAACAAAG 60
DB 1 TGGAGGGGCTTTCACAGGCTCAACCACTGATGCCACTTCTGTCCCAACAAAG 60
QY 61 CAGGAGAGACAACTTCCCTTAAGTGGGTACAGGCTTCTGTGGCTTGGGCT 120
DB 61 CAGGAGAGACAACTTCCCTTAAGTGGGTACAGGCTTCTGTGGCTTGGGCT 120
QY 121 CAGGCCCCCTTCATCATGGGATCAAAATGTGAAAGTGTCTCATACGCTAAAGCTACT 180
DB 121 CAGGCCCCCTTCATCATGGGATCAAAATGTGAAAGTGTCTCATACGCTAAAGCTACT 180
QY 181 CTGCGCGGGCCAAACCCCTTGTGTATAGCTGGAGCCGTCCAAAACGAGGTCAACCTTC 240
DB 181 CTGCGCGGGCCAAACCCCTTGTGTATAGCTGGAGCCGTCCAAAACGAGGTCAACCTTC 240
QY 241 ACACACCCCATTAACCAATTCAATGCGATGATGTCAGCCGACCTGGAGGTCTGCACG 300
DB 241 ACACACCCCATTAACCAATTCAATGCGATGATGTCAGCCGACCTGGAGGTCTGCACG 300
QY 301 AGCACCCTGGGTCTGGTGGGCGGGGCTTGGAGCTCTGGAGCTCTGCTGATTTGTTGCAACA 360
DB 301 AGCACCCTGGGTCTGGTGGGCGGGGCTTGGAGCTCTGGAGCTCTGCTGATTTGTTGCAACA 360
QY 361 GGCACGCTGTGATTTGTGGGTAGATCACTTGTCCGGGCGGCGCTATTGTTCCCGAC 420
DB 361 GGCACGCTGTGATTTGTGGGTAGATCACTTGTCCGGGCGGCGCTATTGTTCCCGAC 420
QY 421 AGGGAAGTCTTACCAAGAGTTCCATGATGATGAAAGATGGCGTTCGACTCCCTTAC 480
DB 421 AGGGAAGTCTTACCAAGAGTTCCATGATGATGAAAGATGGCGTTCGACTCCCTTAC 480
QY 481 ATCGAGAGGGAATGAGCTCGCCGAGCAGTTCAAGCAAAAAGCGCTCGGTTGCTGCAG 540
DB 481 ATCGAGAGGGAATGAGCTCGCCGAGCAGTTCAAGCAAAAAGCGCTCGGTTGCTGCAG 540
QY 541 ACAGCCACCAACCAAGCGAGGCGCGTGTCTCCGTTGGTGAAGTCCAAATGGCGAGCCCTT 600
DB 541 ACAGCCACCAACCAAGCGAGGCGCGTGTCTCCGTTGGTGAAGTCCAAATGGCGAGCCCTT 600
QY 601 GAGACCTTCTGGGCGAAACAATGTGAACTTCAACGCGGATACAGTACTTACGACGCG 660
DB 601 GAGACCTTCTGGGCGAAACAATGTGAACTTCAACGCGGATACAGTACTTACGACGCG 660
QY 661 TTGTCACTCTGCTGGGAATCCCGGATGTCATGATGCGCTTCAAGCTCTGTC 720
DB 661 TTGTCACTCTGCTGGGAATCCCGGATGTCATGATGCGCTTCAAGCTCTGTC 720
QY 721 ACTAGCCGCTCAACCAATCTACCTCTGCTTAAATCATCTGGGGGATGGGTAGCC 780
DB 721 ACTAGCCGCTCAACCAATCTACCTCTGCTTAAATCATCTGGGGGATGGGTAGCC 780
QY 781 GCCCACTGCTCCCCAGTGTGCTTCAAGCTTCTGTAAGGCGCGCAATTGCTGTGGG 840
DB 781 GCCCACTGCTCCCCAGTGTGCTTCAAGCTTCTGTAAGGCGCGCAATTGCTGTGGG 840
QY 841 GGTGTTGGAGATAGGCTTGGGAAGTGTGTTGGAATCATCTGGCGGGCTATGAGAGA 900
DB 841 GGTGTTGGAGATAGGCTTGGGAAGTGTGTTGGAATCATCTGGCGGGCTATGAGAGA 900
QY 901 GGAAGTGCAGGCGCTGCTGAGCTTAAAGTCAATAGCGCGAAATGCTTCAACGAG 960
DB 901 GGAAGTGCAGGCGCTGCTGAGCTTAAAGTCAATAGCGCGAAATGCTTCAACGAG 960
QY 961 GACTGTGTTAACTTACTCCCTGCACTCTCTCTCTGTGGCTTGGTCTGCGGGGTCTGG 1020
DB 961 GACTGTGTTAACTTACTCCCTGCACTCTCTCTCTCTGTGGCTTGGTCTGCGGGGTCTGG 1020

DB 961 GACTGTGTTAACTTACTCCCTGCACTCTCTCTCTCTGTGGCTTGGTCTGCGGGGTCTGG 1020
QY 1021 TGGCAGGAGATACCTGCTCGGACCTGTGGTCCAGGGAGGGGGCTGTCACTGATGAAC 1080
DB 1021 TGGCAGGAGATACCTGCTCGGACCTGTGGTCCAGGGAGGGGGCTGTCACTGATGAAC 1080
QY 1081 CGGCTGATAGGTTTGCTCCGCGGGGTAAACATGTTTCCCCACGCACTATGTGCCAAG 1140
DB 1081 CGGCTGATAGGTTTGCTCCGCGGGGTAAACATGTTTCCCCACGCACTATGTGCCAAG 1140
QY 1141 AGCGAGCGCGAGACGATGTCATGATACCTCCGACTTATCATCAACCAACTGTGG 1200
DB 1141 AGCGAGCGCGAGACGATGTCATGATACCTCCGACTTATCATCAACCAACTGTGG 1200
QY 1201 AAGAGCTCAACAGTGAATTAAACAGAGACTGCTCAACGCTGTCCGACTCGTGGCTA 1260
DB 1201 AAGAGCTCAACAGTGAATTAAACAGAGACTGCTCAACGCTGTCCGACTCGTGGCTA 1260
QY 1261 AGGATGTTTGGAGCTGATATGCAACAGTTTGGCTGACTTCAAGCCTGGCTCCAGTTC 1320
DB 1261 AGGATGTTTGGAGCTGATATGCAACAGTTTGGCTGACTTCAAGCCTGGCTCCAGTTC 1320
QY 1321 AAGCTCTGCGCGGATTAACGGGAGTCCGCTTTTCTCATGCGCAAGTGGTAAAGGGG 1380
DB 1321 AAGCTCTGCGCGGATTAACGGGAGTCCGCTTTTCTCATGCGCAAGTGGTAAAGGGG 1380
QY 1381 GTCTGGCGGGAGACGGGATATGCAACACTGCTCATGTGAGGACAGATCACCGGA 1440
DB 1381 GTCTGGCGGGAGACGGGATATGCAACACTGCTCATGTGAGGACAGATCACCGGA 1440
QY 1441 CATGTCAAAAACGGTTCATGAGATCGTTGGGCTTAAAGCTGTATGTAACATGTGGCAT 1500
DB 1441 CATGTCAAAAACGGTTCATGAGATCGTTGGGCTTAAAGCTGTATGTAACATGTGGCAT 1500
QY 1501 GGAACATTCCTCATCAAGCATCAACACAAGGCGCTGCAAGCCCTCCCAAGCGCAAAAC 1560
DB 1501 GGAACATTCCTCATCAAGCATCAACACAAGGCGCTGCAAGCCCTCCCAAGCGCAAAAC 1560
QY 1561 TATTCAGGCGCTGTGGCGGGTGGCTGTGAGAGTACGTGAGGTTACGCGGTTGGGG 1620
DB 1561 TATTCAGGCGCTGTGGCGGGTGGCTGTGAGAGTACGTGAGGTTACGCGGTTGGGG 1620
QY 1621 GATTTTCACTACGTACAGACATGACACTGACCAACGTAATAATGCCCGTCCAGGTTCCA 1680
DB 1621 GATTTTCACTACGTACAGACATGACACTGACCAACGTAATAATGCCCGTCCAGGTTCCA 1680
QY 1681 GCCCGCGAATCTTCAAGAGTGAATGGGGTGGCGGCTGCAAGGTAAGCTCCGCGTGC 1740
DB 1681 GCCCGCGAATCTTCAAGAGTGAATGGGGTGGCGGCTGCAAGGTAAGCTCCGCGTGC 1740
QY 1741 AAACCTCTCTTACGGAAGAGGTCAATTCAGGTCCGAGCTCAACCAATCTGTGGTGGG 1800
DB 1741 AAACCTCTCTTACGGAAGAGGTCAATTCAGGTCCGAGCTCAACCAATCTGTGGTGGG 1800
QY 1801 TCGAGCTTCCATGCGAGCGGACCGGATGTGAGCATGCTCACTTCCATGCTCACGAC 1860
DB 1801 TCGAGCTTCCATGCGAGCGGACCGGATGTGAGCATGCTCACTTCCATGCTCACGAC 1860
QY 1861 CCTTCCCATACAGAGCGGCTAAAGCGGCAAGCTGGGCCAGGGGGGTCTCCGCCCTCC 1920
DB 1861 CCTTCCCATACAGAGCGGCTAAAGCGGCAAGCTGGGCCAGGGGGGTCTCCGCCCTCC 1920
QY 1921 TTGGCAGGCTCTTACGATGACAGTTGTCTGCGCTTCTCTGAAGGCGCATACATTAAC 1980
DB 1921 TTGGCAGGCTCTTACGATGACAGTTGTCTGCGCTTCTCTGAAGGCGCATACATTAAC 1980
QY 1981 CAAAATGACTTCCAGACGCTGACTCATCGAGCGCAACCTCTGTGGCGGATGAGATG 2040
DB 1981 CAAAATGACTTCCAGACGCTGACTCATCGAGCGCAACCTCTGTGGCGGATGAGATG 2040
QY 2041 GGC 2043
DB 2041 GGC 2043


```

Db      3484 AGGAGTGTGGAGCTGATATGCAAGTTTTGCTGACTTCAAGACCTGGCTCCAGTTC 3543
Oy      1321 AAGCTCTGCGCGATTAACCGGGAGTCCCTTTTCTCATGCGCAAGTGGATCAAGGGG 1380
Db      3544 AAGCTCTGCGCGATTAACCGGGAGTCCCTTTTCTCATGCGCAAGTGGATCAAGGGG 3603
Oy      1381 GTTGGCGGGAGACGGCATCATGCAAGACCACTTGCTCATGTGAGACACATCACCGGA 1440
Db      3604 GTCTGGCGGGAGACGGCATCATGCAAGACCACTTGCTCATGTGAGACACATCACCGGA 3663
Oy      1441 CATGCAAAAACGGTTCCATGAGATCGTTGGGCTTAAGACCTGATAGTAACATGGGCAT 1500
Db      3664 CATGCAAAAACGGTTCCATGAGATCGTTGGGCTTAAGACCTGATAGTAACATGGGCAT 3723
Oy      1501 GGAAATTCCTCCCATACGATACACACACAGGCGCCCTGACCGCTCCCGACGCGCAAC 1560
Db      3724 GGAAATTCCTCCCATACGATACACACAGGCGCCCTGACCGCTCCCGACGCGCAAC 3783
Oy      1561 TATTCCAGGGCGCTGTGGCGGGTGGCTGCTGAGAGTACGTGAGATTACGCGGGTGGGG 1620
Db      3784 TATTCCAGGGCGCTGTGGCGGGTGGCTGCTGAGAGTACGTGAGATTACGCGGGTGGGG 3843
Oy      1621 GATTTCCTACTAGTACACGACGATGACCACTGACCAAGTAAATGCCCGGCGAGGTTCGA 1680
Db      3844 GATTTCCTACTAGTACACGACGATGACCACTGACCAAGTAAATGCCCGGCGAGGTTCGA 3903
Oy      1681 GCCCGCGAATTCCTACAGAAATGATGAGGGGTGGCGGCTGCACAGGTACGCTCCGGCGTGC 1740
Db      3904 GCCCGCGAATTCCTACAGAAATGATGAGGGGTGGCGGCTGCACAGGTACGCTCCGGCGTGC 3963
Oy      1741 AAACCTCTCTACGAGAGAGGTCACTTCCAGGTGCGGCTCAACCAATATCTGTGTTGGG 1800
Db      3964 AAACCTCTCTACGAGAGAGGTCACTTCCAGGTGCGGCTCAACCAATATCTGTGTTGGG 4023
Oy      1801 TGGGAGCTTCCCTGAGAGCGCCGACCGGATGATGACAGTGTCTCACTTCCATGCTCACCGAC 1860
Db      4024 TGGGAGCTTCCCTGAGAGCGCCGACCGGATGATGACAGTGTCTCACTTCCATGCTCACCGAC 4083
Oy      1861 CCTCCCATCATACAGACAGACAGCGGTACCGCAGGCTGGCCAGGGGGTCTCCCGCTCC 1920
Db      4084 CCTCCCATCATACAGACAGACAGCGGTACCGCAGGCTGGCCAGGGGGTCTCCCGCTCC 4143
Oy      1921 TTGGCCAGCTCTTCAAGTACGAGTTGTCTGCGCTTCTCTGAAGGCGACATATATAC 1980
Db      4144 TTGGCCAGCTCTTCAAGTACGAGTTGTCTGCGCTTCTCTGAAGGCGACATATATAC 4203
Oy      1981 CAAATGATTTCCCGACGCTGACCTCATCGAGGCGCAACCTCTGTGGCGGATGAGATG 2040
Db      4204 CAAATGATTTCCCGACGCTGACCTCATCGAGGCGCAACCTCTGTGGCGGATGAGATG 4263
Oy      2041 GGC 2043
Db      4264 GGC 4266

RESULT 3
ACAE1114
ID      ACAE1114 strand; cDNA; 5955 BP.
XX
AC      ACAE1114;
XX
DT      09-JUL-2003 (first entry)
DE      HCV cDNA encoding Met-NS3-NS4A-NS4B-NS5A-NS5B (active).
XX
KM      HCV; ss; gene; non-structural protein; NS3; NS4A; NS4B; NS5A; NS5B;
XX      adenoviral vector; HCV infection; vaccine; gene therapy; protease.
OS      Hepatitis C virus.
XX
FH      Key      Location/Qualifiers
FT      CDS      1..5955
FT

```

```

FT      /product= "Met-NS3-NS4A-NS4B-NS5A-NS5B"
FT      /partial
FT      /note= "No stop codon shown"
XX
XX      W02003031588-A2.
XX
XX      17-APR-2003.
XX
XX      10-OCT-2002; 2002WO-US032512.
XX
XX      11-OCT-2001; 2001US-0328655P.
XX      13-MAR-2002; 2002US-0363774P.
XX
XX      (MERI ) MERCK & CO INC.
XX      (RICE-) IST RICERCHE BIOL MOLECOLARE ANSELETTI.
XX      Emili BA, Kariow DC, Bett AJ, Shiver JW, Nicotia A, Lahm A;
XX      Luzzago A, Cortese R, Colloca S;
XX      WPI; 2003-381708/36.
XX      P-PEDB; AB009575.
XX
XX      New nucleic acid encoding a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide,
XX      useful as a component of an adenovector or DNA plasmid vaccine for
XX      preventing or treating hepatitis C virus.
XX
XX      Example 1; Fig 5; 231pp; English.
XX
XX      The invention relates to a nucleic acid encoding a Met-NS3-NS4A-NS4B-NS5A
XX      -NS5B (NS stands for non-structural protein from HCV) polypeptide
XX      appearing as AB009574. The encoded polypeptide has sufficient protease
XX      activity to process itself to produce an NS5B protein that is
XX      enzymatically inactive. Also included are a cultured recombinant cell
XX      comprising the novel nucleic acid, making an adenovector (comprising: (a)
XX      producing an adenovirus genome plasmid comprising a gene expression
XX      cassette by homologous recombination between the novel nucleic acid and a
XX      nucleic acid comprising a first adenovirus region from base pair 1-450
XX      corresponding to either Ad5 or Ad6, a second adenovirus region from base
XX      pair 3511-5548 corresponding to Ad5 or from base pair 3508-5541
XX      corresponding to Ad6, joined to the first region, a third adenovirus
XX      region from base pair 5549-2813 corresponding to Ad5 or from base pair
XX      5542-28156 corresponding to Ad6, joined to the second region, a fourth
XX      adenovirus region from base pair 30818-33966 corresponding to Ad5 or from
XX      base pair 30789-33784 corresponding to Ad6, joined to the third region,
XX      and a fifth adenovirus region from base pair 33967-35935 corresponding to
XX      Ad5 or from base pair 33785-35759 corresponding to Ad6, joined to the
XX      fourth region; and (b) rescuing the recombinant adenovirus from the
XX      recombinant adenovirus plasmid), an adenoviral vector that is produced by
XX      method above appearing as ACAE1113 which has a humanised version of the
XX      polynucleotide of the invention and encodes the HCV inactivated
XX      polynucleotide, a recombinant nucleic acid comprising one or more Ad6
XX      regions and a region not present in Ad6, where at least one Ad6 region is
XX      selected from E1A, E1B, E2A, E2B, E4, L1, L2, L4 and L5. The nucleic acid
XX      is useful as a component of an adenoviral vector or a DNA plasmid vaccine
XX      providing a broad range of antigens for generating an HCV-specific cell
XX      mediated immune response. The nucleic acid may also be used in treating
XX      patients infected with HCV. The present sequence is the non-humanised
XX      cDNA encoding the wild-type HCV polypeptide with an active NS5B protein
XX
XX      Sequence 5955 BP; 1251 A; 1798 C; 1678 G; 1228 T; 0 U; 0 Other;
XX
XX      Query Match      88.3%; Score 1803; DB 8; Length 5955;
XX      Best Local Similarity 92.7%; Pred. No. 0;
XX      Matches 1893; Conservative 0; Mismatches 150; Indels 0; Gaps 0;
XX
Oy      1 TGGAGGGCGCTTTCAACAGGCTTACACCGGATGAGTGGCCACTTCTGCGCAACCAAG 60
Db      1597 TGGAGAGTGTCTTCAAGGCTTACACCGGATGAGTGGCCACTTCTTGTCCAGACCAAG 1656
Oy      61 CAGGAGAGACCAACTCCCTACTGTTGGGTACAGAGCTACTGTGTGCGCTTACGGGCC 120
Db      1657 CAGGAGAGACCAACTCCCTACTGTTGGGTACATACAGACCAAGCGGTGTGCGCGGCT 1716

```

QY 121 CAGGCCCACTTCATCATGGATCAAAATGAGAGTGTCTATACGAGCTAAAGCCTACT 180
 DB 1717 CAGGCCCACTTCATCATGGATCAAAATGAGAGTGTCTATACGAGCTAAAGCCTACT 1776
 QY 181 CTGCGCGGCGCAACCTTCTGTATAGGCTGGAGCGCTCCAAAACGAGCTCACCTTC 240
 DB 1777 CTGCAACGGGCGCAACCTTCTGTATAGGCTGGAGCGCTCCAAAATAGGCTCACCTTC 1836
 QY 241 ACACACCCCATTAACAAATTCATCATGGATGCAATGTCAGCGGACCTGAGAGGCTGCAG 300
 DB 1837 ACCCACCCCATTAACAAATTCATCATGGATGCAATGTCAGCTGAGAGGCTGCAGCT 1896
 QY 301 AGCACTGGGTGCTGGTGGCGGAGTCTTTCAGAGCTCTGAGCTGCTATTTCTTGACACA 360
 DB 1897 AGCACTGGGTGCTGGTGGCGGAGTCTTTCAGAGCTCTGAGCTGCTATTTCTTGACACA 1956
 QY 361 GGCAGCGTGTCAATGTGGGTAGATCATCTTGTCCGGCGGCGGCTATTTGTTCCCGAC 420
 DB 1957 GGCAGTGTGTATGTGGGTAGATTCCTTGTCCGGGAGGCGGCTATTTGTTCCCGAC 2016
 QY 421 AGGGAAGTCTCTACAGAGATTCATGATGGAAGAGTGGGCGTGGACCTCCCTTAC 480
 DB 2017 AGGGAATTCCTACAGAGATTCATGATGGAAGAGTGGGCGTGGACCTCCCTTAC 2076
 QY 481 ATCGACAGGGAATGAGCTCGCGGAGAGTTCAGACAAAAGCGCTCGGTTTCTGACAG 540
 DB 2077 ATCGAGAGGGAATGAGCTCGCGGAGAGTTCAGACAAAAGCGCTCGGTTTCTGACAG 2136
 QY 541 ACAGCCACCAAGCAAGCGGAGCGCTGCTCCGCTGGTGGAGTCCAAATGGCGAGCCTT 600
 DB 2137 ACAGCCACCAAGCAAGCGGAGCGCTGCTCCGCTGGTGGAGTCCAAATGGCGAGCCTT 2196
 QY 601 GAGACCTTCTGGGCGAAACATATGTGAACTTCATACAGGAGGATAAGTACTTAAAGAGCG 660
 DB 2197 GAGACCTTCTGGGCGAAACATATGTGAACTTCATACAGGAGGATAAGTACTTAAAGAGCG 2256
 QY 661 TTGTCACTCTGCTGGGAAATCCCGGATGTGATCATCTAGGCGGCTTCAAGCCTGTATC 720
 DB 2257 TTATCACTCTGCTGGGAAATCCCGGATGTGATCATCTAGGCGGCTTCAAGCCTGTATC 2316
 QY 721 ACTAGCCGCTCACAACCAATCTACCTCTGCTTAAACATCTGGGGGAGATGAGTACCC 780
 DB 2317 ACTAGCCGCTCACAACCAATCTACCTCTGCTTAAACATCTGGGGGAGATGAGTACCC 2376
 QY 781 GCCCAATCTGCTCCCCCGAGTCTGTTAGGCTTTGTAGGCGCGGCAATGTGTGTGG 840
 DB 2377 GCCCAATCTGCTCCCCCGAGTCTGTTAGGCTTTGTAGGCGCGGCAATGTGTGTGG 2436
 QY 841 GCTGTGGCAGATAGGCTTGGGAAAGTGTGTGGAATCTTGGCGGAGTATGAGAGA 900
 DB 2437 GCTGTGGCAGATAGGCTTGGGAAAGTGTGTGGAATCTTGGCGGAGTATGAGAGA 2496
 QY 901 GGAAGTGGAGGCGGCTGTGGCTTTAAAGTATAGAGCGGCAATGCTTCCACGAG 960
 DB 2497 GGAAGTGGAGGCGGCTGTGGCTTTAAAGTATAGAGCGGCAATGCTTCCACGAG 2556
 QY 961 GACCTGTGTAATCTTACTCTCTGTCATCTCTCTGCTGTGGTGTCTGGGAGTGTGG 1020
 DB 2557 GACCTGTGTAATCTTACTCTCTGTCATCTCTCTGCTGTGGTGTCTGGGAGTGTGG 2616
 QY 1021 TGCAGAGCATATCTGTGCGGACGATGGGTCCAGGGGAGGGAGCTGTGACATGATGAAC 1080
 DB 2617 TGCAGAGCATATCTGTGCGGACGATGGGTCCAGGGGAGGGAGCTGTGACATGATGAAC 2676
 QY 1081 CGGCTGATAGCGTTCGCTCGCGGGGTAAACATGTTTCCCGCAGCAGTATGTGCCAG 1140
 DB 2677 CGGCTGATAGCGTTCGCTCGCGGGGTAAACATGTTTCCCGCAGCAGTATGTGCCAG 2736
 QY 1141 AGCGAGCGCGGACAGATGTCTACTGATCTCTCCGACCTTACTATCAACCACTGTGG 1200
 DB 2737 AGCGAGCGCGGACAGATGTCTACTGATCTCTCCGACCTTACTATCAACCACTGTGG 2796
 QY 1201 AAGAGCTCCACCAAGTATTAACGAGAGCTGTCCAGCCCTGCTCCGGCTGTGGCTTA 1260

DB 2797 AAAAGGCTCCACAGTGGATTAATGAAGACTGCTCCACACCGGTGTTCCGGCTGTGGCTTA 2856
 QY 1261 AGGAGTGTGGGACCTGATATAGCAACATTTTGGCTGACTTCAAGACCTGGCTCCAGTCC 1320
 DB 2857 AGGAGTGTGGGACCTGATATAGCAACATTTTGGCTGACTTCAAGACCTGGCTCCAGTCC 2916
 QY 1321 AAGCTCTCCCGGATTAACCGGAGTCCCTTTTCTCATGTGCAACGTGGGTACAAGGG 1380
 DB 2917 AAGCTCTCCCGGATTAACCGGAGTCCCTTTTCTCATGTGCAACGTGGGTACAAGGG 2976
 QY 1381 GTCTGGCGGGAGAGCGGATTCATGCAACCACTGCTCATGTGAGCACAGATCACCGGA 1440
 DB 2977 GTCTGGCGGGAGAGCGGATTCATGCAACCACTGCTCATGTGAGCACAGATCACCGGA 3036
 QY 1441 CATGTCAAAAAGGTTCCATGAGATTCGTGGGCGCTTAAGACCTGTATTAACATGTGGCAT 1500
 DB 3037 CATGTCAAAAAGGTTCCATGAGATTCGTGGGCGCTTAAGACCTGTATTAACATGTGGCAT 3096
 QY 1501 GGAACATTCGCCATCAACGATACACACAGGAGCCCTGACGCGCTCCCAAGCGCCAAC 1560
 DB 3097 GGAACATTCGCCATCAACGATACACACAGGAGCCCTGACACACCTCTCCAGCGCCAAC 3156
 QY 1561 TATTCAGGGGCTGTGTGGGCTGTGTGAGAGTACGTGAGGTACGCGGCTGGG 1620
 DB 3157 TATTCAGGGGCTGTGTGGGCTGTGTGAGAGTACGTGAGGTACGCGGCTGGG 3216
 QY 1621 GATTTCCATATGATGACGAGCATGACCAATGACAACTGAAATGCCCGGCGCAGTTCCA 1680
 DB 3217 GATTTCCATATGATGACGAGCATGACCAATGACAACTGAAATGCCCGGCGCAGTTCCA 3276
 QY 1681 GCCCGGAAATTCCTCAACAGATGATGAGGTGGGCTGCGGCTGACAGGATGCTCCGGGCTGC 1740
 DB 3277 GCTCTGAAATTCCTCAACAGATGATGAGGTGGGCTGCGGCTGACAGGATGCTCCGGGCTGC 3336
 QY 1741 AAACCTCTCTCAACGAGAGAGTCAATTCAGGTGCGGCTCAACCAATACCTGTTGG 1800
 DB 3337 AGGCTCTCTCAACGAGAGAGTCAATTCAGGTGCGGCTCAACCAATACCTGTTGG 3396
 QY 1801 TGCAGCTCCATGGAAGCGGCAACCGGATGTAGAGTGTCTCAATGTCTCAACCGAC 1860
 DB 3397 TGCAGCTCCATGGAAGCGGCAACCGGATGTAGAGTGTCTCAATGTCTCAACCGAC 3456
 QY 1861 CCTCCCATCATCAACGAGAGAGCGGTAAGCGGAGCTGCGGAGGAGTCTCCCGCTCC 1920
 DB 3457 CCTCCCATCATCAACGAGAGAGCGGTAAGCGGAGCTGCGGAGGAGTCTCCCGCTCC 3516
 QY 1921 TTGGCAGCTCTTCAAGCTAGCCAGTGTGTGCGCTTCTCTCAAGAGCGCATATCATTTAC 1980
 DB 3517 TTGGCAGCTCTTCAAGCTAGCCAGTGTGTGCGCTTCTCTCAAGAGCGCATATCATTTAC 3576
 QY 1981 CAATATGACTTCCAGAGCGGTGACTCATGAGGCGCAACCTCTGTGGCGGCAATGAGT 2040
 DB 3577 CAATATGACTTCCAGAGCGGTGACTCATGAGGCGCAACCTCTGTGGCGGCAATGAGT 3636
 QY 2041 GGC 2043
 DB 3637 GGC 3639

RESULT 4
 ID ACA61111 standard, cDNA, 5965 BP.
 XX ACA61111;
 AC ACA61111;
 XX
 DT 09-JUL-2003 (first entry)
 XX
 DE HCV cDNA encoding Met-NS3-NS4A-NS4B-NS5A-NS5B (inactivated).
 HCV, ss; gene; non-structural protein; NS3; NS4A; NS4B; NS5A; NS5B;
 KM adenoviral vector; HCV infection; vaccine; mutant; gene therapy;
 protease.

| Query Match | Best Local Similarity | 88.3% | Score 1803; | DB 8; | Length 5965; |
|----------------------------|-----------------------|--------------|-----------------|-----------|--------------|
| Matches 1993; Conservative | 92.7%; | Pred. No. 0; | Mismatches 150; | Indels 0; | Gaps 0 |

| | | | | | | |
|----|------|-------------------|----------------------------|----------------------------------|--------------------|------|
| QY | I | TGGGAGGGCGCTCTTCA | CAGGCGCTTCA | CCCAAGTGAGATGCCCACTTCCTGTC | CCCAACAAG | 60 |
| Db | 1603 | TGGGAGAGTGCTTTCA | CAGGCGCTCA | CCCAATAGATCAACTTCCTGTC | CCAGACAAG | 1662 |
| QY | 61 | CAGCAGAGAGCAACTT | CCCCCTACCTGAGGCGGACA | CAGGCTACTGTCGCGATAGGGCC | 120 | |
| Db | 1663 | CAGCAGAGAGCAACTT | CCCCCTACCTGAGATCCAA | GCACGGGTGTGTGCGCAGGGCT | 1722 | |
| QY | 121 | CAGGCCCCACCTCCAT | CATGGGATCAAA | TGTGTGAAAGTGTCTCA | TACGGCTTAAAGCTTACT | 180 |
| Db | 1723 | CAGGCCCCACCTCCAT | CATGGGATCAAA | TGTGTGAAAGTGTCTCA | TACGGCTTAAAGCTTACT | 1782 |
| QY | 181 | CTGCGCGGGCGCA | CACTTGTCTGTAT | AGCTGTGGAGCCGCTCCAAA | ACAGAGTCAACCTTC | 240 |
| Db | 1783 | CTGCACGCGGCGCA | CACTTGTCTGTAT | AGAGCTGTGGAGCCGCTCCAAA | TGAGGTCAACCTTC | 1842 |
| QY | 241 | ACACACCCCAATAC | CAAAATTCATATGG | CAATGCAATGACCCGACCTGTGAGAGT | CGTACAG | 300 |
| Db | 1843 | ACCAACCCCAATAC | CAAAATTCATATGG | CAATGCAATGACCTGTGAGAGT | CGTACACT | 1902 |
| QY | 301 | AGCACTTGTGTCGT | TGTGGGCGGGGTCCTTG | CAGCTCTGCTGTGATTTGCTTGACA | CAACA | 360 |
| Db | 1903 | AGCACTTGTGTCGT | TGTGGGCGGGGTCCTTG | CAGCTCTGCTGTGATTTGCTTGACA | CAACA | 1962 |
| QY | 361 | GGCAGCGGTGAT | TGTTGGGTAGATCAT | CTTGTCCGGGCGGCGCGCTATTTGTTCCCGAC | 420 | |
| Db | 1963 | GGCAGGTGTGAT | TGTTGGGTAGATCAT | CTTGTCCGGGCGGCGCGCTATTTGTTCCCGAC | 2022 | |
| QY | 421 | AGGGAAGTCCCTTA | CCAGAGTTGAGATG | AGATGAGATGCGAGTGGCACTCCCTTAC | 480 | |
| Db | 2023 | AGGGAAGTTCCTTA | CCAGAGTTGAGATG | AGATGAGATGCGAGTGGCACTCCCTTAC | 2082 | |
| QY | 481 | ATCAGACAGGAAAT | GACAGCTCGCCGACAGTTCA | AGCAAAAAGGCTTCGAGTTGCTTCAG | 540 | |
| Db | 2083 | ATCAGACAGGAAAT | GACAGCTCGCCGACAGTTCA | AGCAAAAAGGCTTCGAGTTGCTTCAG | 2142 | |
| QY | 541 | ACAGCCACCAAG | CGAAGCGGAGCGCTGCTCCGCTGTG | AGATTCAAATGTGCGACGCTT | 600 | |
| Db | 2143 | ACAGCCACCAAG | CGAAGCGGAGCGCTGCTCCGCTGTG | AGATTCAAATGTGCGAGCTT | 2202 | |
| QY | 601 | GAGACCTTCTGGG | GGAACAATGTGAACTTCA | TCAAGCGGAGATACAGTACTTATGCA | AGGC | 660 |
| Db | 2203 | GAGACCTTCTGGG | GGAACAATGTGAAATTCATCA | AGCGGAGATACAGTACTTATGCA | AGGC | 2262 |
| QY | 661 | TTGTCAACTCTG | CGTGGGAAATCCCGCATTTG | CACTCATGTATGCGGTTCA | CACAGCTCTGTC | 720 |
| Db | 2263 | TTATTCACACTG | CGCTGGGAACCCCGCAATAG | CACTCATGTATGGAATTCACAGCTCTATC | 2322 | |
| QY | 721 | ACTAGCCCGCTAC | CAACCCAAATCTAACCTCTGCTTAA | CATCTGTGGGGATGTGGTATGCC | 780 | |
| Db | 2323 | ACAGACCCGCTAC | CAACCCAAAGTACCTCTCTGTTTAA | CACTTGTGGGGGGTGTGGTATGCC | 2382 | |
| QY | 781 | GCCCAACTCTG | TCCCCCGCAGTCTGCTTCA | AGCGCGCGGCATTTGCTGTGTGCG | 840 | |
| Db | 2383 | GCCCAACTCTG | TCCCCCGCAGCGCTTCTGCTTTC | GTGTGTGCGCGCGCATGTCCGCGTGTG | 2442 | |
| QY | 841 | GCTGTGTGACAT | ATAGGCTTGTGGAGATGCTTGTG | ACAATCTTGAGCGCGGCTATGAGACA | 900 | |
| Db | 2443 | GCTGTGTGACAT | ATAGGCTTGTGGAGATGCTTGTG | ACAATCTTGCGCGGGTATGAGACA | 2502 | |
| QY | 901 | GGAAGTGCAGG | CGGCGCTGTGTGCTTTA | AGGTCAATGAGCGCGCAATGCCCTCCAC | CGAG | 960 |
| Db | 2503 | GGAAGTGCAGG | CGGCGCTGTGTGCTTTA | AGGTCAATGAGCGCGCAATGCCCTCCAC | CGAG | 2562 |
| QY | 961 | GACCTGTGTTAA | CTTACTCTCTGCAATCTCTCTCTGTG | CGCTGCTGTGTCTGTGTGTGGGGTGTG | 1020 | |
| Db | 2563 | GACCTGTGTTAA | CTTACTCTCTGCAATCTCTCTCTGTG | CGCTGCTGTGTGTGTGTGTGGGGTGTG | 2622 | |
| QY | 1021 | TGCGAGAGCAAT | ACGCGTCCGCAAGTGGGATC | AGAGGAGGAGGAGCTGTGCACTGATGAC | 1080 | |
| Db | 2623 | TGTGAGCAAAAT | ACGCGTCCGCAAGTGGGATC | AGAGGAGGAGGAGCTGTGCACTGATGAC | 2682 | |
| QY | 1081 | CGGCTGATAG | GTTGCTCGCTCGCGGGGTAA | CCATGTTCCTCCACGCACTATGTGCA | AG | 1140 |

```

Db      2683  CCGGCTGATAGCCTTGGCCCTCGCGGGGTATCATGTTCCGCCACGCACTATGCTGAG 2742
Oy      1141  AGCGAGCGCGGAGACGCTGCTCACTGAGATCTCTCCGACCTTACTATCAACCACTGTTG 1200
Db      2743  AGCGAGCGCGGAGACGCTGCTCACTGAGATCTCTCCGACCTTACTATCAACCACTGTTG 2802
Oy      1201  AAGAGGCTTCAACGATGATTAACGAGAGCTGCTCCAGCCCTGCTCCGGCTGCTGAGCTA 1260
Db      2803  AAAAGGCTTCAACGATGATTAACGAGAGCTGCTCCAGCCCTGCTCCGGCTGCTGAGCTA 2862
Oy      1261  AGGAGTGTGGGACTGATATGACAGATTTTGGCTGACTTCAAGACCTGGCTCCAGTCC 1320
Db      2863  AGGAGTGTGGGACTGATATGACAGATTTTGGCTGACTTCAAGACCTGGCTCCAGTCC 2922
Oy      1321  AAGCTTCGCGCGGCAATACCGGGAGTCCCTTTTCTCATGCGCAAGTGGGTAACAAGGG 1380
Db      2923  AAGCTTCGCGCGGCAATACCGGGAGTCCCTTTTCTCATGCGCGCAAGTGGGTAACAAG 2982
Oy      1381  GTCTGGCGGGAGAGCGGATCATGACGACCACTGCTCATGTGAGACAGATCAACGGGA 1440
Db      2983  GTCTGGCGGGAGAGCGGATCATGACGACCACTGCTCATGTGAGACAGATCAACGGGA 3042
Oy      1441  CATGTCAAAAACGGTTCATGAGAGATGTTGGGCTTAAGACCTGTATGTAACATGTGGCAT 1500
Db      3043  CATGTCAAAAACGGTTCATGAGAGATGTTGGGCTTAAGACCTGTATGTAACATGTGGCAT 3102
Oy      1501  GGAACATTTCCCATCAACGATACACACAGGAGGCTTGCAGCGCTTCCCGACGCGCAAC 1560
Db      3103  GGAACATTTCCCATCAACGATACACACAGGAGGCTTGCAGCGCTTCCCGACGCGCAAC 3162
Oy      1561  TATTCAGAGGCGCTGAGGCGGCTGCTGAGAGATGCTGAGAGGTTACGCGGGGTGGG 1620
Db      3163  TATTCAGAGGCGCTGAGGCGGCTGCTGAGAGATGCTGAGAGGTTACGCGGGGTGGG 3222
Oy      1621  GATTTCCACTAGTGAACGAGATGACACTGACCAACGTAATAATGCCCGTGCAGATTCCA 1680
Db      3223  GATTTCCACTAGTGAACGAGATGACACTGACCAACGTAATAATGCCCGTGCAGATTCCA 3282
Oy      1681  GCGCCCGAATTTTTCACAGAAATGATGGAGTGGCGCTGCACAGATACCTCCGGGTGC 1740
Db      3283  GCTCTGAATTTTTCACAGAAATGATGGAGTGGCGCTGCACAGATACCTCCGGGTGC 3342
Oy      1741  AAACCTCTCTAGGAGGAGGATTCATTCAGATGCGGCTCAACCAATACCTGTTGGG 1800
Db      3343  AGGCTCTCTCTAGGAGGAGGATTCATTCAGATGCGGCTCAACCAATACCTGTTGGG 3402
Oy      1801  TCGAGCTTCCATGCGAGGCGCAACCGAGTGTAGACAGTGTCTCATGTCTCAACGAC 1860
Db      3403  TCGAGCTTCCATGCGAGGCGCAACCGAGTGTAGACAGTGTCTCATGTCTCAACGAC 3462
Oy      1861  CCTTCCCATCATACAGACGAGACCGGTAAGCGGAGGCTGCGCAAGGGGGTCTCCCTTC 1920
Db      3463  CCTTCCCATCATACAGACGAGACCGGTAAGCGGAGGCTGCGCAAGGGGGTCTCCCTTC 3522
Oy      1921  TTGGCAGCTTCTTCACTAGCAGATTTGTGGGCTTCTCGAAGGCGCATTAATTACC 1980
Db      3523  TTGGCAGCTTCTTCACTAGCAGATTTGTGGGCTTCTCGAAGGCGCATTAATTACC 3582
Oy      1981  CAAAATGATTTCCAGACGCTGACCTCATGAGGCGCAACCTCTGTGGCGGATGAGATG 2040
Db      3583  CACCATGATTTCCAGACGCTGACCTCATGAGGCGCAACCTCTGTGGCGGATGAGATG 3642
Oy      2041  GGC 2043
Db      3643  GGC 3645

```

RESULT 5
 ID ACA61113 standard; DNA; 37090 BP.
 XX
 AC ACA61113;

```

XX      09-JUL-2003 (first entry)
DT      Adenovirus vector MRKAD6-NSmut.
XX      HCV, day non-structural protein; NS3, NS4A, NS4B, NS5A, NS5B;
DE      adenoviral vector; HCV infection; vaccine; gene therapy; protease.
XX      Hepatitis C virus.
XX      Adenovirus serotype 5.
OS      Adenovirus serotype 6.
OS      Synthetic.
XX      Key      Location/Qualifiers
FT      CDS      1264..7221
FT      /tag= a
FT      /product= "Met-NS3-NS4A-NS4B-NS5A-NS5B (inactive)"
XX      WO2003031588-A2.
XX      17-APR-2003.
XX      10-OCT-2002; 2002MO-US032512.
XX      11-OCT-2001; 2001US-0328655P.
XX      13-MAR-2002; 2002US-0363774P.
XX      (MERI ) MERCK & CO INC.
XX      (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.
XX      Eminl EA, Kaslow DC, Bett AJ, Shiver JW, Nicolsia A, Lahm A;
XX      Luzzago A, Cortese R, Colloca S;
XX      WPI: 2003-381708/36.
XX      P-PSDB; AB009574.
XX      New nucleic acid encoding a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide,
XX      useful as a component of an adenovector or DNA plasmid vaccine for
XX      preventing or treating hepatitis C virus.
XX      Claim 32; Fig 4; 231bp; English.
XX      The invention relates to a nucleic acid encoding a Met-NS3-NS4A-NS4B-NS5A
XX      -NS5B (NS stands for non-structural protein from HCV) polypeptide
XX      appearing as AB009574. The encoded polypeptide has sufficient protease
XX      activity to process itself to produce an NS5B protein that is
XX      enzymatically inactive. Also included are a cultured recombinant cell
XX      comprising the novel nucleic acid, making an adenovector (comprising: (a)
XX      producing an adenovirus genome plasmid comprising a gene expression
XX      cassette by homologous recombination between the novel nucleic acid and a
XX      nucleic acid comprising a first adenovirus region from base pair 1-450
XX      corresponding to either Ad5 or Ad6, a second adenovirus region from base
XX      pair 3511-5548 corresponding to Ad5 or from base pair 3508-5541
XX      corresponding to Ad6, joined to the first region, a third adenovirus
XX      region from base pair 5549-28133 corresponding to Ad5 or from base pair
XX      5542-28156 corresponding to Ad6, joined to the second region, a fourth
XX      adenovirus region from base pair 30818-33966 corresponding to Ad5 or from
XX      base pair 30789-33784 corresponding to Ad6, joined to the third region,
XX      and a fifth adenovirus region from base pair 33967-35935 corresponding to
XX      Ad5 or from base pair 33785-35759 corresponding to Ad6, joined to the
XX      fourth region; and (b) rescuing the recombinant adenovirus from the
XX      recombinant adenovirus plasmid), an adenoviral vector that is produced by
XX      method above appearing as ACA61113 which has a humanised version of the
XX      polynucleotide of the invention and encodes the HCV inactivated
XX      polypeptide, a recombinant nucleic acid comprising one or more Ad6
XX      regions and a region not present in Ad6, where at least one Ad6 region is
XX      selected from E1A, E1B, E2A, E2B, E4, E4, E4, E4, E4 and E5. The nucleic acid
XX      is useful as a component of an adenoviral vector or a DNA plasmid vaccine
XX      providing a broad range of antigens for generating an HCV-specific cell
XX      mediated immune response. The nucleic acid may also be used in treating
XX      patients infected with HCV. The present sequence is an adenoviral vector
XX      containing Ad5 and Ad6 sequences together with an HCV promoter and the
XX      non-humanised cDNA encoding the HCV polypeptide with an inactive NS5B

```

| CC | protein | Sequence | 37090 BF; 8382 A; 10762 C; 10299 G; 7647 T; 0 U; 0 Other; |
|----|---------|----------------------------|---|
| XX | | | |
| SO | | | |
| | | Query Match | 88.3%; Score 1803; DB 8; Length 37090; |
| | | Best Local Similarity | 92.7%; Pred. No. 0; |
| | | Matches 1893; Conservative | 0; Mismatches 150; Indels 0; Gaps 0; |
| QY | | 1 | TGGAGGGGCGTCCTTACAGGCGCTCACGCCAGCTGATGCGCCACTTCTGTGCCAACAAG 60 |
| DB | | 2860 | TGGAGAGATGTCCTTACAGGCGCTCACGCCAGCTGATGCGCCACTTGTGCCAACAAG 2919 |
| QY | | 61 | CAGGACAGAGAACAACTTCCCTTACTCTGGTGGCGATCCAGGCTACTGTGTGGCTAAGGGCC 120 |
| DB | | 2920 | CAGGACAGAGAACAACTTCCCTTACTCTGGTGGCGATCCAGGCTACTGTGTGGCTAAGGGCT 2979 |
| QY | | 121 | CAGGCGCCACCTCCATCATGAGGATCAATGTGGAAGTGTCTCATACGGCTTAAAGCTACT 180 |
| DB | | 2980 | CAGGCGCCACCTCCATCATGAGGATCAATGTGGAAGTGTCTCATACGGCTTAAAGCTACT 3039 |
| QY | | 181 | CTGGCGGGGCAACACCCCTTGTCTGTATAGGCTGGAGGCGCTCCAAAAGAGTCAACCTC 240 |
| DB | | 3040 | CTGGACGGGGCAACACCCCTTGTCTGTATAGGCTGGAGGCGCTCCAAAAGAGTCAACCTC 3099 |
| QY | | 241 | ACACACCCCATTAACCAATTATCATATGAGCATATGTACGCCAGCTGAGAGTGTCTACG 300 |
| DB | | 3100 | ACCCACCCCATTAACCAAAATATCATATGAGCATATGTGTGGCTGACCTGAGAGTGTCTACT 3159 |
| QY | | 301 | AGCACCTGGGTGCTGTGTGGGGGGGCTCTTGAGAGCTCGGTGTGAGTATTTGTTGACAA 360 |
| DB | | 3160 | AGCACCTGGGTGCTGTGTGGGGGGGCTCTTGAGAGCTCTGGCGCGGTATTTGCTGACAA 3219 |
| QY | | 361 | GGCAGCGTGTATATTGTGGGTAGATCATCTTGTCCGGGCGGCGCGCTATTGTTCCCGAC 420 |
| DB | | 3220 | GGCAGGTGTGTATATTGTGGGTAGATCATCTTGTCCGGGAGGCGCGCTATTGTTCCCGAC 3279 |
| QY | | 421 | AGGGAAGTCTCTTACACAGAGTTGCATGATGAAAGATGCGGTGCGACCTCCCTTAC 480 |
| DB | | 3280 | AGGGAGTTCTCTTACACAGAGTTGCATGAAATGGAAGATGCGGTGCGACCTCCCTTAC 3339 |
| QY | | 481 | ATCGAGCAGGGGAATGACGCTGCGCGAGAGTTCAAGCAAAAAGGCTGCGGTTGCTGAG 540 |
| DB | | 3340 | ATCGAGCAGGGGAATGACGCTGCGCGAGAGTTCAAGCAAAAAGGCTGCGGTTGCTGCA 3399 |
| QY | | 541 | ACAGCACCAAGCAAGCGAGGCGCTGCTCCCGTGTGAGATCCAAATGCGAGCCCTT 600 |
| DB | | 3400 | ACAGCACCAAGCAAGCGAGGCGCTGCTCCCGTGTGAGATCCAAATGCGAGCCCTT 3459 |
| QY | | 601 | GAGACCTTCTGGGCGAACAACATGTGAACTTCATCAGCGGGATACGATCTTAAGCAGGC 660 |
| DB | | 3460 | GAGACCTTCTGGGCGAACAACATGTGGAATTCATCAGCGGGATACGATCTTAAGCAGGC 3519 |
| QY | | 661 | TTGTGCATCTGCGCGTGGGAATCCGCGGATTCAGATGAGGTGCGTTACAGCCCTGTGC 720 |
| DB | | 3520 | TTATTCATCTGCGCGTGGGAATCCGCGGAATGATATTGATGAGCATTCACAGCCCTATTC 3579 |
| QY | | 721 | ACTAGCCCGCTACCAACCAATCTAACCTCTGCTTAAATCATCTGAGGAGATGGGTAGCC 780 |
| DB | | 3580 | ACCAACCCCGCTACCAACCAAGTAACTCTCTGTTTAACATCTTGGGGGGGTGGGTGCT 3639 |
| QY | | 781 | GCCCACTTCGCTCCCCCGCAGTGTCTTCACTTTCGTAAGGCGCGGCAATTTGTGTGTCG 840 |
| DB | | 3640 | GCCCAACTTCGCTCCCCCGCAGTGTCTTTCGAGGCGCGGCAATTCGCGCGGTGTCG 3699 |
| QY | | 841 | GCTGTTGGCAGATGAGCTTGGGAAGTGTCTTGTGGAATCTTGGCGGGCTAATGAGACA 900 |
| DB | | 3700 | GCTGTTGGCAGATGAGCTTGGGAAGTGTCTTGTGGAATCTTGTGGCGGGTTAATGAGACA 3759 |
| QY | | 901 | GGAGTGGCAGAGCGGCTCGTGGCTTTTAAAGTCATGACGGCGGAATCCCTCCACGAG 960 |
| DB | | 3760 | GGAGTGGCAGAGCGGCTCGTGGCTTTTCAAGGTCAATGAGCGGAGATGCCCTCCACGAG 3819 |
| QY | | 961 | GACCTGGTTAACTTACTCCCTGACATCTCTCTCTGTGTCCTGGTGTGTGTCGGGGTGTG 1020 |

| | | | |
|----|------|---|------|
| Db | 3820 | GACCTGGTCAATCTACTCTTCCTGCGATCTCTCTCTGCGCCCTGGTCTGCTGGGGTCTGT | 3879 |
| Qy | 1021 | TGCGCAGCGATCTACTGCGTCCGACGTGGGTTCAGGGGAGGGGCTGTGCAGTGGATGAAC | 1080 |
| Db | 3880 | TGTGACAGCAATACTGCGTCCGACAGTGGGTCCGGAGAGGGGGCTGTGCAGTGGATGAAC | 3939 |
| Qy | 1081 | CGGCTGATAGGTTTCGGCTCGCGGGGTAAACAATGTTTCCCGACGTAAATGTGCCAGAG | 1140 |
| Db | 3940 | CGGCTGATAGGTTTCGGCTCGCGGGGTAAACAATGTTTCCCGACGTAAATGTGCCAGAG | 3999 |
| Qy | 1141 | AGCGACGCGCAGCAGCGTGCACCTCAGATCCTCTCCGACCTTACTATACCCAACTGTG | 1200 |
| Db | 4000 | AGCGACGCGCAGCAGCGTGTACTCAATCTCTCCAGCCTTACATCACTCAGCTGCTG | 4059 |
| Qy | 1201 | AAGAGGCTCCACAGTGGATTTAACGAGAGCTGCTCAAGCCCTGCTCCGCTGTGGCTTA | 1260 |
| Db | 4060 | AAAAGGCTCCACAGTGGATTTAATGAGAGCTGCTCCACAGGTTGTCGGCTGTGGCTTA | 4119 |
| Qy | 1261 | AGGAGTGTTTGGAGCTGGATATGACAGCTTTGGCTCACTTCAAGACCTGGCTCCAGTCC | 1320 |
| Db | 4120 | AGGAGTGTTTGGAGCTGGATATGACAGGTTGTGACTACTTCAAGACTTGGCTCCAGTTC | 4179 |
| Qy | 1321 | AAGCTCCGCGCGCATTAACCGGAGTCCCTTTTCTCATGSCCAAGTGGGTACAAGGG | 1380 |
| Db | 4180 | AAGCTCTGCGCAGCTACCGGGAGTCCCTTTTCTCTGTGSCCAAGCGGGGTACAGGGA | 4239 |
| Qy | 1381 | GTCTGGCGGGGAGACGGGATCATGACAGACCACTGCTCATGTGGAGACAGATCACCGGA | 1440 |
| Db | 4240 | GTCTGGCGGGGAGACGGGATCATGCAAAACCACTGCCCATGTGGAGACAGATCACCGGA | 4299 |
| Qy | 1441 | CATGTCAAAAAAGTTTCATATAGGATGTGTTGGGCTTAAGAAGCTGTAACTAATATGTGGCAT | 1500 |
| Db | 4300 | CATGTCAAAAAAGTTTCATATAGGATGTGTGGGCTTAAGAGCTGTACACAAACAAGTGCAT | 4359 |
| Qy | 1501 | GGAACTATCCCATCAACAGCATACACACGAGGCCCTCGACAGCCCTCCGACGSCAAAC | 1560 |
| Db | 4360 | GGAACTATCCCATCAACAGCATACACAGGAGGCCCTCGACACCCCTTCCAGCGCCAAAC | 4419 |
| Qy | 1561 | TATTCAGAGGCGCTGTGGCGGGTGGCTGTGAGAGTAACTGTGAAGTTAACGGGGTGGG | 1620 |
| Db | 4420 | TATTCAGAGGCGCTGTGGCGGGTGGCGGTGAGAGTAACTGTGAAGTAAACGGGGTGGG | 4479 |
| Qy | 1621 | GATTTCCACTACGTGACGAGCATACCACTGACAACTGAATGCCGTGCAGGTTTCCA | 1680 |
| Db | 4480 | GATTTCCACTACGTGACGAGCATACCACTGACAACTGAATGCCCATGTGCAGGTTTCCG | 4539 |
| Qy | 1681 | GCCCCCGAATCTTCAACAAGTGGAGTGGGGTGGGGCGACAGGTAAGCTCCGGCGTGC | 1740 |
| Db | 4540 | GCTCTGAATCTTCAACAAGTGGAGTGGGGTGGGGCGACAGGTAAGCTCCGGCGTGC | 4599 |
| Qy | 1741 | AAACCTCTCTACCGGAGAGAGSTACATTCAGGTCGGGCTCAACCAATACCTGTGTGG | 1800 |
| Db | 4600 | AGGCTCTCTCTACCGGAGAGAGSTTAACTTCAAGTCCGGGCTCAACCAATACCTGTGTGG | 4659 |
| Qy | 1801 | TGCGAGCTCCCATGCGAGCCCGAAACCGAGTGTACAGTGTCACTTCAATGTCTACCGAC | 1860 |
| Db | 4660 | TGCGAGCTCCCATGCGAGCCCGAAACCGAGTGTACAGTGTCACTTCAATGTCTACCGAC | 4719 |
| Qy | 1861 | CCCTCCCACTACACAGACGAGACGGCTTAAGGCGAGGCTGGCAGGGGGTCTCCGCCCTCC | 1920 |
| Db | 4720 | CCCTCCCACTACACAGACGAGAAAGGCTTAAGGTTGGCAGGGGGTCTCCGCCCTCC | 4779 |
| Qy | 1921 | TTGGCAGACTCTTCACTAGCAGCAATGTGTCTGCGCTTCTTGAAGGGGACATACATTACC | 1980 |
| Db | 4780 | TTGGCAGACTCTTCACTAGCAGCAATGTGTCTGCGCTTCTTGAAGGGGACATACATTACC | 4839 |
| Qy | 1981 | CAAAATGAATTCAGACGCTGACTTCAATGAGGSCAACTCTGTGTGGCGGCATGAGATG | 2040 |
| Db | 4840 | CACATATGCTCTCGGAGCGTGAAGCTCATGAGGCCAACTCTGTGTGGCGGACAGATG | 4899 |
| Qy | 2041 | GGC 2043 | |
| | | | |

DB 4900 GGC 4902

RESULT 6
ID ADR38452 Standard; DNA; 7987 BP.
XX ADR38452;
XX 18-NOV-2004 (first entry)
XX
XX DNA encoding Hepatitis C virus (isolate BK) NS3-NS5B polypeptide Seq 3.
XX
XX Hepatitis C virus; HCV; NS3; NS5A; non-structural protein;
XX
XX virus replication activity; gene; ds; mutant.
XX
XX Hepatitis C virus (isolate BK).
XX
XX Synthetic.
XX
XX
XX Key Location/Qualifiers
XX 5'UTR 1..386
XX /tag= a
XX CDS 387..1181
XX /tag= b
XX /product= "Beta-lactamase"
XX misc_feature 1225..1800
XX /tag= c
XX /label= EMCV IRES
XX /note= "internal ribosome entry site"
XX CDS 1801..7758
XX /tag= d
XX /product= "NS3-NS5B polypeptide"
XX /gene= "NS3 (1804-3696)"
XX /gene= "NS4A (3697-3858)"
XX /gene= "NS4B (3859-4641)"
XX /gene= "NS5A (4642-5982)"
XX /gene= "NS5B (5983-7755)"
XX 3'UTR 7759..7987
XX /tag= e
XX
XX WO2004074507-A2.
XX
XX 02-SEP-2004.
XX
XX 09-FEB-2004; 2004WO-US003726.
XX
XX 13-FEB-2003; 2003US-0447318P.
XX
XX (MERI) MERCK & CO INC.
XX
XX Grobler J, Flores O, Markel EJ;
XX
XX MPI; 2004-635590/61.
XX P-PSDB; ADR38450.
XX
XX
XX Making Hepatitis C virus (HCV) replicon having increased replication
XX activity, useful in HCV research, comprises modifying HCV replicon
XX construct to encode an amino acid substitution at a position
XX corresponding to amino acid 470 of NS3.
XX
XX Claim 10; SEQ ID NO 3; 54pp; English.
XX
XX This invention relates to a novel method for producing a Hepatitis C
XX virus (HCV) replicon having an increased replication activity.
XX Specifically, it refers to modifying an HCV replicon construct to encode
XX an amino acid substitution in NS3 (a non-structural protein that along
XX with NS4A, NS4B, NS5A and NS5B make up the virus replication machinery
XX released in the form of a polypeptide). The present invention describes
XX an amino acid substitution at a position corresponding to amino acid 470
XX of NS3 alone, or in combination with, an isoleucine in a position
XX corresponding to amino acid 232 of NS5A that confers improved cell
XX culture replication activity compared to wild type HCV. The method is
XX useful for facilitating the identification of broadly efficacious

CC compounds against different HCV isolates and facilitating HCV research,
CC where compounds that inhibit HCV replication have research and
CC therapeutic applications in identifying overall efficacy and lack of
CC unacceptable toxicity. Accordingly, they can be used to treat or inhibit
CC the onset of HCV in a patient. This polynucleotide is a mutant HCV
CC replicon DNA sequence that encodes an HCV NS3-NS5B polypeptide (based on
CC HCV-BK) with 471M and 1179I residue substitutions, given in an
CC exemplification of the invention.
XX
XX Sequence 7987 BP; 1722 A; 2315 C; 2204 G; 1746 T; 0 U; 0 Other;
XX
XX Query Match 88.2%; Score 1801.4; DB 13; Length 7987;
XX Best Local Similarity 92.6%; Pred. No. 0;
XX Matches 1892; Conservative 0; Mismatches 151; Indels 0; Gaps 0;
XX
QY 1 TGGGAGGGGCTTTCACAGGCTCCACCGGATGAGGCTTCTGCTCCCAACAAAG 60
DB 3397 TGGGAGAGTGTCTTACAGGCTCCACCGATGAGGCTTCTGCTCCCAACAAAG 3456
QY 61 CAGGAGAGAGCACTTCCCTACCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 120
DB 3457 CAGGAGAGAGCACTTCCCTACCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 3516
QY 121 CAGGAGAGAGCACTTCCCTACCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 180
DB 3517 CAGGAGAGAGCACTTCCCTACCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 3576
QY 181 CTGGCCGCGGCGCAACCTTGTGTATGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 240
DB 3577 CTGGCCGCGGCGCAACCTTGTGTATGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 3636
QY 241 ACACAGCCCATACCAATTCATCATGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 300
DB 3637 ACACAGCCCATACCAATTCATCATGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 3696
QY 301 AGCAGCTGGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 360
DB 3697 AGCAGCTGGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 3756
QY 361 GGCAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 420
DB 3757 GGCAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 3816
QY 421 AGGAGAGCTTCTTACAGGAGTTCATGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 480
DB 3817 AGGAGAGCTTCTTACAGGAGTTCATGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 3876
QY 481 ATGAGAGAGGAGTTCATGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 540
DB 3877 ATGAGAGAGGAGTTCATGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 3936
QY 541 ACAGCCACCAAGCAAGGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 600
DB 3937 ACAGCCACCAAGCAAGGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 3996
QY 601 GAGACCTTGGGCGGAAACATGAGGCTTCAAGGAGGCTTCAAGGAGGCTTCAAGGAGGCT 660
DB 3997 GAGACCTTGGGCGGAGGAGGCTTCAAGGAGGCTTCAAGGAGGCTTCAAGGAGGCT 4056
QY 661 TTATCCACTCTGCTGAGGAGTCCGAGTTCATGAGGCTTCAAGGAGGCTTCAAGGAGGCT 720
DB 4057 TTATCCACTCTGCTGAGGAGTCCGAGTTCATGAGGCTTCAAGGAGGCTTCAAGGAGGCT 4116
QY 721 ACTAGCCGCTACCAACCAATTCATGAGGCTTCAAGGAGGCTTCAAGGAGGCTTCAAGGAGGCT 780
DB 4117 ACTAGCCGCTACCAACCAATTCATGAGGCTTCAAGGAGGCTTCAAGGAGGCTTCAAGGAGGCT 4176
QY 781 GCCCACTGCTGCTGAGGAGTCCGAGTTCATGAGGCTTCAAGGAGGCTTCAAGGAGGCT 840
DB 4177 GCCCACTGCTGCTGAGGAGTCCGAGTTCATGAGGCTTCAAGGAGGCTTCAAGGAGGCT 4236
QY 841 GCTGTGAGGAGTTCATGAGGCTTCAAGGAGGCTTCAAGGAGGCTTCAAGGAGGCTTCAAGGAGGCT 900

| | | | |
|----|------|--|------|
| Db | 4237 | GCTGTGGCAGCATAGGCTTGGGAAGGTCTTGTGGACATTCTGGCGGGTTATGAGCA | 4238 |
| Qy | 901 | GGAGGAGCAGGCGCGCTCTGGCCCTTTAAAGTCATGACGGCGAAATGCCCTCCACCGAG | 960 |
| Db | 4297 | GGAGTGGCCGGCGCGCTGTGGCCCTTCAAGGTCATGACGGCGAGATGCCCTCCACCGAG | 4356 |
| Qy | 961 | GACCTGGTTAACTTACTCCCTGGCATTCTCTCTCTGGTGGCCCTGGTGGTGGGGTGGT | 1020 |
| Db | 4357 | GACCTGGTTAACTTACTCTCTGGCATTCTCTCTCTGGCCTGGTGGTGGGGGTGGT | 4416 |
| Qy | 1021 | TGCGGAGGATATCTGTCGGGACGTGGGATCCAGGGGAGGGGGGCTGTCACTGATGTAAC | 1080 |
| Db | 4417 | TGTGAGAGCAATATCTGTCGACGACTGGGGTCCGGGAGGGGGGCTGTGCACTGATGTAAC | 4476 |
| Qy | 1081 | CGGCTGATAGCGTTGGCCTCGCGGGGGTAAACATGTTTCCCGACGCACTATGTGCCAAG | 1140 |
| Db | 4477 | CGGCTGATAGCGTTGGCCTCGCGGGGGTAAATATGTTTCCCGACGCACTATGTGTGAG | 4536 |
| Qy | 1141 | AGCGACGGCGGACGACGTTGCACTAGATCTCTCCGACTTTACTATACCCAACTGTTG | 1200 |
| Db | 4537 | AGCGACGGCGGACGCGCTGTGTTACTAGATCTCTCCGACTTACATCACTAGCTGTG | 4596 |
| Qy | 1201 | AAGAGGCTCCACCAATGGAATTAAGGAGGACTGCTCAGGCCGTGCTCGGCTCGTGCTA | 1260 |
| Db | 4597 | AAAGAGCTCCACCAATGGAATTAAGAGCTGCTCACAACGTTGCTCGGCTCGTGCTA | 4656 |
| Qy | 1261 | AGGAGTGTGGGAGTGTGATATGACAGATTTTGGGTGACTTCAAGACCTGGCTCAATCC | 1320 |
| Db | 4657 | AGGAGTGTGGGAGTGTGATATGACAGGTTGTTGACTGACTTCAAGACCTGGCTCAATCC | 4716 |
| Qy | 1321 | AAGCTCTGCGCGCATTAACCGGAGTCCCTTTTTCATGCAACGTGGGTAACAAGGG | 1380 |
| Db | 4717 | AAGCTCTGCGCGCACTAACCGGAGTCCCTTTTTCGTGTCACAAAGGGGTAACAAGGA | 4776 |
| Qy | 1381 | GTCGTGGCGGGGAGAGGGATCATGAGACACACTGCTCATGTGGAGGACAAATCACGGGA | 1440 |
| Db | 4777 | GTCGTGGCGGGGAGAGGGATCATGTGAACACACTGCTCATGTGGAGGACAAATCACGGGA | 4836 |
| Qy | 1441 | CATGTCAAAAACGTTCCATGACGATCGTTGGGCTTAAGACCTGTGATGATGTGGCAT | 1500 |
| Db | 4837 | CATGTCAAAAACGTTCCATGAGATCGTGGGCTTAAGACCTGTGAGAACACGTGGCAT | 4896 |
| Qy | 1501 | GGAACATTCCCATCAACGACATACACACGGGCCCTGACGCGCTCCCAAGCGCAAC | 1560 |
| Db | 4897 | GGAACATTCCCATCAACGACATACACACGGGCCCTGACACCTCTCCACAGCGCAAC | 4956 |
| Qy | 1561 | TATTCACAGGCGCTGTGGCGGGTGGCTGTAGAGATACGTGAGATTAACCGGGTGGGG | 1620 |
| Db | 4957 | TATTCACAGGCGCTGTGGCGGGTGGCTGTAGAGATACGTGAGATTAACCGGGTGGGG | 5016 |
| Qy | 1621 | GATTTCCACTACGTACGACGATGACCACTGACCAACGTAAAAATGCGCTGCAGGTTCCA | 1680 |
| Db | 5017 | GATTTCCACTACGTACGAGGATGACCACTGACCAACGTAAAAATGCGCTGCAGGTTCCG | 5076 |
| Qy | 1681 | GCCCCGAATTTCTTCAAGAAATGAGTGGGGTGGCTGCAACGATACGTTCCGGCGTGC | 1740 |
| Db | 5077 | GCTTCCGAATTTCTTCAAGGAGGTGACGAGTGGCGGTTGCAACGATACGTTCCGGCGTGC | 5136 |
| Qy | 1741 | AAACCTTCTTACGGGAGGAGGTCACTTCCAGGTCGGGCTTCAACCACTACTGTTGGG | 1800 |
| Db | 5137 | AGGCTTCTTCAACGGGAGGAGGTCACTTCCAGGTCGGGCTTCAACCACTACTGTTGGG | 5186 |
| Qy | 1801 | TGCGAGCTCCCATGTGGAGCCGAAACGGATGTGAGAGATGTCATCTTCATGCTACCGAC | 1866 |
| Db | 5197 | TGCGAGCTTCCCATGTGGAGCCGAAACGGATGTGAGAGATGTCATCTTCATGCTACCGAC | 5256 |
| Qy | 1861 | CCCTTCCCATCAACGACAGACGCGCTAAGCGCAGGCTGCGCACGGGGGTCTTCCCTTCC | 1920 |
| Db | 5257 | CCCTTCCCATCAACGACAGAAACGGCTAAGGTGAGGTTGGCCACGGGGGTCTTCCCTTCC | 5316 |
| Qy | 1921 | TTGGCCAGCTCTTCAAGCTAGCCAGTTGTCTGGCGCTTCTCTGAGAGGCAATACCTAAC | 1980 |
| Db | 5317 | TTGGCCAGCTCTTCAAGCTATCAGTTGTCTGGCGCTTCTCTTGAAGGCAATGCACTTACC | 5376 |

| | | | |
|----------|--|---|------|
| QY | 1961 | CAAAATGACTTCCCAAGACGCTGACCTCATCGAGGCCAACTCTCTGTGGCCGCAAGATG | 204 |
| Db | 5377 | CACCATGTCTCTCCGAGCGCTGACCTCATCGAGGCCAACTCTCTGTGGCCGCAAGATG | 5436 |
| QY | 2041 | GGC 2043 | |
| | | | |
| Db | 5437 | GGC 5439 | |
| RESULT 7 | | | |
| ADNR2189 | | | |
| ID | ADNR2189 standard; DNA; 9587 BP. | | |
| XX | | | |
| AC | ADNR2189; | | |
| XX | | | |
| DT | 16-DEC-2004 (first entry) | | |
| XX | | | |
| DE | Hepatitis C virus type 1b polyprotein DNA. | | |
| XX | | | |
| KM | antlihemie; cardiant; vasotropic; antiarteriosclerotic; antidiabetic; | | |
| KM | cytostatic; anticonvulsant; nootropic; muscular; anti-HIV; | | |
| KM | RNA interference; RNA; antisense technology; lipid metabolism; | | |
| KM | cholesterol imbalance; dyslipidaemia hypercholesterolaemia; | | |
| KM | coronary artery disease; CAD; coronary heart disease; CHD; | | |
| KM | atherosclerosis; hepatic glucose production; | | |
| KM | glucose-metabolism-related disorder; diabetes; cancer; breast cancer; | | |
| KM | colon cancer; lung cancer; neurological disease; Huntington disease; | | |
| XX | spinocerebellar ataxia; viral disease; AIDS; HCV; polyprotein; gene; ds. | | |
| OS | Hepatitis C virus. | | |
| XX | | | |
| PN | WO2004080406-A2. | | |
| XX | | | |
| PD | 23-SEP-2004. | | |
| XX | | | |
| PF | 08-MAR-2004; 2004WO-US007070. | | |
| XX | | | |
| PR | 07-MAR-2003; 2003US-0452682P. | | |
| PR | 12-MAR-2003; 2003US-0454265P. | | |
| PR | 13-MAR-2003; 2003US-0454962P. | | |
| PR | 13-MAR-2003; 2003US-0455050P. | | |
| PR | 14-APR-2003; 2003US-0462894P. | | |
| PR | 17-APR-2003; 2003US-0463772P. | | |
| PR | 25-APR-2003; 2003US-0465665P. | | |
| PR | 25-APR-2003; 2003US-0465802P. | | |
| PR | 09-MAY-2003; 2003US-0469612P. | | |
| PR | 08-AUG-2003; 2003US-0493986P. | | |
| PR | 11-AUG-2003; 2003US-0494597P. | | |
| PR | 26-SEP-2003; 2003US-0506341P. | | |
| PR | 09-OCT-2003; 2003US-0510246P. | | |
| PR | 10-OCT-2003; 2003US-0510318P. | | |
| PR | 07-NOV-2003; 2003US-0518453P. | | |
| PA | (ALNY-) ALNYTAM PHARM. | | |
| XX | | | |
| PI | Manoharan M, Bumcrot D; | | |
| XX | | | |
| DR | WPI, 2004-677362/66. | | |
| PT | | | |
| PT | Interference RNA agent useful for treating dyslipidaemias, coronary artery | | |
| CC | disease, diabetes, cancer or neurological disease, comprises sense | | |
| CC | sequence and antisense sequence which has specific modifications. | | |
| XX | | | |
| PS | Example 5; SEQ ID NO 6688, 378pp; English. | | |
| XX | | | |
| CC | The invention describes a RNA interference (iRNA) agent (I) comprising a | | |
| CC | sense sequence and an antisense sequence, where the sense sequences have | | |
| CC | one or more asymmetrical 2'-O alkyl modifications, the antisense | | |
| CC | sequences have one or more asymmetrical phosphorochiostate modifications | | |
| CC | and the antisense sequence targets a human gene sequence. Also described | | |
| CC | are the pharmaceutical preparation comprising (I); reducing (M1) apob-100 | | |
| CC | levels or glucose-6-phosphatase levels in a subject; producing (I); | | |

CC stabilising (1), involves selecting a sequence with activity and
 CC introducing one or more asymmetrical modification in the sequence, where
 CC the modification decreases nuclease sensitivity while not decreasing its
 CC activity; a kit comprising (I) and instruction for its use; and a device
 CC that can be dispense or administer a composition comprising (I). (I) is
 CC useful for reducing apob-100 levels or glucose-6-phosphatase levels. (M1)
 CC is useful for reducing apob-100 levels or glucose-6-phosphatase levels.
 CC The subject is suffering from a disorder characterised by elevated or
 CC otherwise unwanted expression of apob-100, elevated or otherwise unwanted
 CC levels of cholesterol, and/or dysregulation of lipid metabolism. The
 CC disorder is chosen from the HDL/LDL cholesterol imbalance,
 CC hyperlipidaemia, hypercholesterolaemia, statin-resistant
 CC dyslipidaemia, coronary artery disease (CAD), coronary heart
 CC disease (CHD) and atherosclerosis. (I) is administered to a subject to
 CC inhibit hepatic glucose production or for treating glucose-metabolism-
 CC related disorder e.g. diabetes or type-2 diabetes. (I) is useful for
 CC treating the diseases as mentioned above, cancer (e.g. breast, colon or
 CC lung cancer), neurological disease (e.g., Huntington disease or
 CC spinocerebellar ataxia) or viral disease (e.g., AIDS). This sequence
 CC represents hepatitis C virus type 1b polypeptide DNA.

XX Sequence 9587 BP, 1921 A, 2865 C, 2714 G, 2087 T, 0 U, 0 Other;

Query Match 88.2%; Score 1801.4; DB 13; Length 9587;
 Best Local Similarity 92.6%; Pred. No. 0;
 Matches 1892; Conservative 0; Mismatches 151; Indels 0; Gaps 0;

QY 1 TGGGAGGGGCTTTTCAACAGGCTTCAACCCAGTGGATGCCACTTCTGCTCCAAACAAAG 60
 Db 5013 TGGGAGAGTGTCTTCAACAGGCTTCAACCCAGTGGATGCCACTTCTGCTCCAAACAAAG 5072
 QY 61 CAGGAGAGAGCAACTTCCCTACCTGAGGAGTACAGGCTACTGAGGCTAGGAGGCT 120
 Db 5073 CAGGAGAGAGCAACTTCCCTACCTGAGGAGTACAGGCTACTGAGGCTAGGAGGCT 5132
 QY 121 CAGGCCCCACTTCATCATGAGATCAAAATGTGAAAGTGTCTCATACGCTTAAAGCTTACT 180
 Db 5133 CAGGCCCCACTTCATCATGAGATCAAAATGTGAAAGTGTCTCATACGCTTAAAGCTTACT 5192
 QY 181 CTGGCGGGGGCCAAACCTTGTCTGATAGGCTGGAGAGCTGCAAAACAGATCACCTTC 240
 Db 5193 CTGGCGGGGGCCAAACCTTGTCTGATAGGCTGGAGAGCTGCAAAACAGATCACCTTC 5252
 QY 241 ACACACCCCATTAACAAATTCATCATGAGATCATGATGATGAGGCTGAGGAGGCTGACG 300
 Db 5253 ACACACCCCATTAACAAATTCATCATGAGATCATGATGATGAGGCTGAGGAGGCTGACG 5312
 QY 301 AGCACCTGGGTGCTGATGGGCGGGGTCTTGCAGCTTGGCTGCTGATTTGTTGACAAAC 360
 Db 5313 AGCACCTGGGTGCTGATGGGCGGGGTCTTGCAGCTTGGCTGCTGATTTGTTGACAAAC 5372
 QY 361 GGCAGCGGTGATTTGTTGATGATCATTTGTCGGGGCGGCGCTATTTGTTCCGAC 420
 Db 5373 GGCAGCGGTGATTTGTTGATGATCATTTGTCGGGGCGGCGCTATTTGTTCCGAC 5432
 QY 421 AGGGAAGTCTCTACAGAGATTCATGATGATGATGATGATGATGATGATGATGATGATGAT 480
 Db 5433 AGGGAAGTCTCTACAGAGATTCATGATGATGATGATGATGATGATGATGATGATGATGAT 5492
 QY 481 ATCGAGAGGAGATCAGCTCCGAGAGAGTCAAGCAAAAGAGCTCGGGTTGCTGAG 540
 Db 5493 ATCGAGAGGAGATCAGCTCCGAGAGAGTCAAGCAAAAGAGCTCGGGTTGCTGAG 5552
 QY 541 ACAGCCACCAAGCAAGCGAGGCGCTGCTCCGAGTGGAGTCCAGTGGCGAGGCTT 600
 Db 5553 ACAGCCACCAAGCAAGCGAGGCGCTGCTCCGAGTGGAGTCCAGTGGCGAGGCTT 5612
 QY 601 GAGACCTTCTGGGGGAAAACATGTGAACTTCATCAGCGGGATTCAGTACTTGAAGGC 660
 Db 5613 GAGACCTTCTGGGGGAAAACATGTGAACTTCATCAGCGGGATTCAGTACTTGAAGGC 5672
 QY 661 TTGTCACTCTGCTGGAGATCCGCGATGTGATCATGATGAGGCTTCAAGCCTCTGTC 720

Db 5673 TTATCCACTCTGCTGGGAAACCCCGCAATGATCATTTGATGAGCATTCACAGCCTCTATC 5732
 QY 721 ACTAGCCGCTACACCAACCAATCTACCTCTGTTAATCATCTGGGGGAGTGGATGCC 780
 Db 5733 ACTAGCCGCTACACCAACCAATCTACCTCTGTTAATCATCTGGGGGAGTGGATGCC 5792
 QY 781 GCCCACTGCTCCCGCAGTGTCTTCACTTCTGATGAGGCGCGGCAATTCGCTGGTGC 840
 Db 5793 GCCCACTGCTCCCGCAGTGTCTTCACTTCTGATGAGGCGCGGCAATTCGCTGGTGC 5852
 QY 841 GCTGTTGGCAGATAGGCTTGGGAGGCTTGGACATTTGGCGGGCTATGAGACA 900
 Db 5853 GCTGTTGGCAGATAGGCTTGGGAGGCTTGGACATTTGGCGGGCTATGAGACA 5912
 QY 901 GGAGTGGAGAGGCGGCTGTGGCTTTAAGTCAAGAGCGGGAATGCCCTCCACGAG 960
 Db 5913 GGAGTGGAGAGGCGGCTGTGGCTTTAAGTCAAGAGCGGGAATGCCCTCCACGAG 5972
 QY 961 GACCTGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1020
 Db 5973 GACCTGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 6032
 QY 1021 TGGCAGCGATATCTGCTGCGACGTTGGGTCAGGGGAGGGGCTGTGCAATGATGAAC 1080
 Db 6033 TGGCAGCGATATCTGCTGCGACGTTGGGTCAGGGGAGGGGCTGTGCAATGATGAAC 6092
 QY 1081 CGGCTGATAGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCT 1140
 Db 6093 CGGCTGATAGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCT 6152
 QY 1141 AGCGAGCGCGAGCAGTGTCTCAGATTCCTCTCCGACTTATCATCAACCAACTGTTG 1200
 Db 6153 AGCGAGCGCGAGCAGTGTCTCAGATTCCTCTCCGACTTATCATCAACCAACTGTTG 6212
 QY 1201 AAGAGCTTCCACCAATGATTAACAGAGACTGCTCACGCGCTGCTCGGCTGTGCTA 1260
 Db 6213 AAGAGCTTCCACCAATGATTAACAGAGACTGCTCACGCGCTGCTCGGCTGTGCTA 6272
 QY 1261 AAGGATTTTGGGATCTGATATGACAGTTTGGCTGACTTCAACCTGGCTCCAGTCC 1320
 Db 6273 AAGGATTTTGGGATCTGATATGACAGTTTGGCTGACTTCAACCTGGCTCCAGTCC 6332
 QY 1321 AAGCTCTGCGCGATTAACCGGAGTCCCTTTTCTCATGCAACGATGGATCAAGGG 1380
 Db 6333 AAGCTCTGCGCGATTAACCGGAGTCCCTTTTCTCATGCAACGATGGATCAAGGG 6392
 QY 1381 GTTGGCGGGGAGACGGGATCATGACACACTGCTCATGTGAGACAAGATCACCGGA 1440
 Db 6393 GTTGGCGGGGAGACGGGATCATGACACACTGCTCATGTGAGACAAGATCACCGGA 6452
 QY 1441 CATGTCAAAAACGGTTCCATGAGATGTTGGGCTTAAGACTGTGATGATCATGTGGCAT 1500
 Db 6453 CATGTCAAAAACGGTTCCATGAGATGTTGGGCTTAAGACTGTGATGATCATGTGGCAT 6512
 QY 1501 GGAAATTTCCCATCAACGATCAACACAGGAGCCCTGCAAGCCTTCCACAGCGCAAC 1560
 Db 6513 GGAAATTTCCCATCAACGATCAACACAGGAGCCCTGCAAGCCTTCCACAGCGCAAC 6572
 QY 1561 TATTCAGGGGCTGTGGCGGGTGTCTGAGAGATCAGTGAAGTAAACGGGGTGGGG 1620
 Db 6573 TATTCAGGGGCTGTGGCGGGTGTCTGAGAGATCAGTGAAGTAAACGGGGTGGGG 6632
 QY 1621 GATTTTCCATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1680
 Db 6633 GATTTTCCATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6692
 QY 1681 GCGCCGAAATTTCTTCAAGAGTGGATGGGGGCGGCTGCAAGGATGAGCTCCGCGTGC 1740
 Db 6693 GCGCCGAAATTTCTTCAAGAGTGGATGGGGGCGGCTGCAAGGATGAGCTCCGCGTGC 6752
 QY 1741 AAACCTCTCTTCAAGGAGAGATCATTCAGATGGGCTCAACCAATTAATCTGTTGGG 1800
 Db 6753 AAACCTCTCTTCAAGGAGAGATCATTCAGATGGGCTCAACCAATTAATCTGTTGGG 6812


```

Db      5424  AGGAGACTTCTCTACAGAGATTGCAATGAAAGAGTGCCTCGACCTCCCTTAC 5483
Qy      481  ATCGAGGAGGGAATGAGCTCGCCGAGCACTTCAAGCAAAAAGCGTCCGGTTGCTGCAG 540
Db      5484  ATCGAGGAGGGAATGAGCTCGCCGAGCACTTCAAGCAAAAAGCGTCCGGTTGCTGCAG 5543
Qy      541  ACAGCCACCAAGCAAGCGGAGCGCTGCTCCGTTGAGTGCATGAGTGCAGAGCTT 600
Db      5544  ACAGCCACCAAGCAAGCGGAGCGCTGCTCCGTTGAGTGCATGAGTGCAGAGCTT 5603
Qy      601  GAGACCTTCTGAGGCAAAACATGTTGAACTTTCATCAGCGGGAATACAGTACTTACAGAGC 660
Db      5604  GAGACCTTCTGAGGCAAAACATGTTGAACTTTCATCAGCGGGAATACAGTACTTACAGAGC 5663
Qy      661  TTGTTCACCTCTGCTGAGGAAATCCCGGATTTGATCATGAGCGCTTACAGCTCTGTC 720
Db      5664  TTATTCACCTCTGCTGAGGAAATCCCGGATTTGATCATGAGCGCTTACAGCTCTTATC 5723
Qy      721  ACTAGCCGCTCACCACCAATCTACCTCTGCTTAACTATCCTGAGGAGGATGGGTAGGC 780
Db      5724  ACCAGCCGCTCACCACCAATCTACCTCTGCTTAACTATCCTGAGGAGGATGGGTAGGC 5783
Qy      781  GCCCACTGCTCCCCCAGTCTGCTTCAAGCTTTGTAAGCGCGCATTTGCTGCTGCG 840
Db      5784  GCCCACTGCTCCCCCAGTCTGCTTCAAGCTTTGTAAGCGCGCATTTGCTGCTGCG 5843
Qy      841  GCTGTGGGAGCATAGGCTTTGGGAAAGTGTCTTGTGAGATCTTGGCGGCTATGAGAGA 900
Db      5844  GCTGTGGGAGCATAGGCTTTGGGAAAGTGTCTTGTGAGATCTTGGCGGCTATGAGAGA 5903
Qy      901  GGAAGTGGAGGAGCGCTGCTGCTTAAAGTCAATGAGGCGGAAATGCCCTCCACCGAG 960
Db      5904  GGAAGTGGAGGAGCGCTGCTGCTTAAAGTCAATGAGGCGGAAATGCCCTCCACCGAG 5963
Qy      961  GACCTGTGTTAACTTACTCCCTGCAATCTCTCTCTGCTGCTGCTGCTGCTGCTGCTG 1020
Db      5964  GACCTGTGTTAACTTACTCCCTGCAATCTCTCTCTGCTGCTGCTGCTGCTGCTGCTG 6023
Qy      1021  TGGCGAGCATTACTGCTGCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080
Db      6024  TGTGAGCAATTACTGCTGCTGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6083
Qy      1081  CGGCTGATAGGCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140
Db      6084  CGGCTGATAGGCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6143
Qy      1141  AGCGAGCGCGGAGCAAGTGTCACTGATCTCTCCGACTTACTATCAACCAACTGTTG 1200
Db      6144  AGCGAGCGCGGAGCAAGTGTCTGATCTCTCCGACTTACTATCAACCAACTGTTG 6203
Qy      1201  AAGAGGCTTCCAGTGTGATTAAGAGAGTGTCTCCAGCTGCTGCTGCTGCTGCTGCTG 1260
Db      6204  AAGAGGCTTCCAGTGTGATTAAGAGAGTGTCTCCAGCTGCTGCTGCTGCTGCTGCTG 6263
Qy      1261  AGGGAATGTTGGGAGCTGATATGCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320
Db      6264  AGGGAATGTTGGGAGCTGATATGCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6323
Qy      1321  AAGCTCTGCTCGGATTAACCGGAGTCCCTTTTCTCAATGCAAGTGTGCTGCTGCTGCTG 1380
Db      6324  AAGCTCTGCTCGGATTAACCGGAGTCCCTTTTCTCAATGCAAGTGTGCTGCTGCTGCTG 6383
Qy      1381  GTCTGGCGGAGAGAGGATCATGAGACCACTGCTCATGTGAGGAGACAGATCACCGGA 1440
Db      6384  GTCTGGCGGAGAGAGGATCATGAGACCACTGCTCATGTGAGGAGACAGATCACCGGA 6443
Qy      1441  CATGTCAAAAAGGTTTCATGAGGATGTTGGGCTTAAGACCTGTAACATGTGGCAT 1500
Db      6444  CATGTCAAAAAGGTTTCATGAGGATGTTGGGCTTAAGACCTGTAACATGTGGCAT 6503
Qy      1501  GGAACTTCCCATCAAGCATACACAGGAGCCCTGACAGCCCTCCCAAGCGCCCAAC 1560

```

```

Db      6504  GGAACTTCCCATCAAGCATACACAGGAGCCCTGACACCCCTTCCAGCCCAAC 6563
Qy      1561  TATTCAGGAGCGCTGTGGCGGCTGCTGCTGAGAGTATGAGGATTACGGCGGATGGG 1620
Db      6564  TATTCAGGAGCGCTGTGGCGGCTGCTGCTGAGAGTATGAGGATTACGGCGGATGGG 6623
Qy      1621  GATTTCCACTAGTACAGAGCATGACCACTGCAACAGTAAATGCCCCTGCCAGGTTCCA 1680
Db      6624  GATTTCCACTAGTACAGAGCATGACCACTGCAACAGTAAATGCCCCTGCCAGGTTCCG 6683
Qy      1681  GCCCCGAATTTTTCACAGAAATGATGAGGAGTGCAGTGCACAGTACGCTCCGCGTGC 1740
Db      6684  GCTTCTGAAATTTCTTCGAGAGGTGACAGAGTGCAGAGTGCACAGGATACGCTCCGCTGC 6743
Qy      1741  AAACCTTCTTACGAGAGAGGTCAATTTCCAGTCCGAGTCCCAACCAATACCTGATGGG 1800
Db      6744  AAGCCTTCTTACGAGAGAGGTCAATTTCCAGTCCGAGTCCCAACCAATACCTGATGGG 6803
Qy      1801  TCGAGCTTCCATGCGAGGCGGCAACCGGATGAGAGTGCATCTTCCATGCTACCGAC 1860
Db      6804  TCGAGCTTCCATGCGAGGCGGCAACCGGATGAGAGTGCATCTTCCATGCTACCGAC 6863
Qy      1861  CCTTCCCATATACAGCAGAGAGCGCTTAAAGCGCAGGCTGCGCAGGAGGCTTCCCTTCC 1920
Db      6864  CCTTCCCATATACAGCAGAGAGCGCTTAAAGCGCAGGCTGCGCAGGAGGCTTCCCTTCC 6923
Qy      1921  TTGGCAGCTTTCAGTACAGGATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1980
Db      6924  TTGGCAGCTTTCAGTACAGGATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6983
Qy      1981  CAAATGACTTCCAGAGCGCTGACCTCATCGAGGCAACCTTCTGCTGCTGCTGCTGCTG 2040
Db      6984  CACCATGTCTTCCGAGCGCTACCTCATCGAGGCAACCTTCTGCTGCTGCTGCTGCTGCTG 7043
Qy      2041  GGC 2043
Db      7044  GGC 7046

RESULT 9
AAQ21829
ID  AAQ21829 standard; cDNA; 9416 BP.
XX
AC  AAQ21829;
XX
XX
DT  25-MAR-2003 (revised)
DT  01-MAY-1992 (first entry)
XX
XX
DE  Non-A, non-B viral genome.
XX
XX
XX  NANBV; vaccine; immunodiagnosis; antigen; antibody; ds.
OS  Non-A.
OS  Non-B hepatitis virus.
XX
XX
FH  Key
FT  CDS
FT  333..677
FT  /tag= a
FT  /product= "C"
FT  /note= "core protein"
FT  678..905
FT  /tag= b
FT  /product= "M"
FT  /note= "matrix protein"
FT  906..1499
FT  /tag= c
FT  /product= "E"
FT  /note= "envelope protein"
FT  1500..2519
FT  /tag= d
FT  /product= "NS1"
FT  2520..3350
FT  /tag= e
FT  CDS

```

```

FT FT /product= "NS2"
FT CDS 3351..5177
FT /tag= f
FT /product= "NS3"
FT CDS 5178..5918
FT /tag= g
FT /product= "NS4a"
FT CDS 5919..6371
FT /tag= h
FT /product= "NS4b"
FT CDS 6372..9365
FT /tag= i
FT /product= "NS5"
XX EP464287-A.
XX PN
XX PD
XX 08-JAN-1992.
XX 28-DEC-1990; 90BP-00314371.
XX PF
XX 25-JUN-1990; 90JP-00167466.
XX PR 31-AUG-1990; 90JP-00230921.
XX PR 09-NOV-1990; 90JP-00305605.
XX PR 17-JUN-1991; 91BP-00401604.
XX PA
XX (OSAU ) UNIV OSAKA.
XX DR WPI; 1992-009617/02.
XX P-PsDB; AAR20091.
XX PT New DNA from non-A, non-B hepatitis virus - and derived antigenic
XX polypeptide(s) useful for diagnostics, blood screening and in vaccines.
XX
XX Claim 1; Fig 2; 89pp; English.
XX
CC The sequence was obtd. from several overlapping "BK" cDNA clones obtd. by
CC "gene walking" using a cDNA clone isolated from a library prepd. from
CC NANBV RNA. The DNA and fragments of it can be used for the detection of
CC the presence of NANBV by hybridisation or PCR. Antigenic polypeptides
CC encoded by the sequence can be used as immunoassay reagents, for
CC screening donated blood, and as immuno- genes for vaccine prodn.
CC Antibodies raised to the peptides can be used in immunoassays to detect
CC or quantify NANBV antigens in liver tissue and blood. Preferred
CC polypeptides are encoded by the following nucleotides: 333-422, -677, or
CC -631; 474-563; 678-905; 906-953 or -1499; 1020-1046 or -1121; 1194-1232;
CC 1209-1322; 1500-2519; 2520-3350; 3351-5177; 4485-4574; 5178-5918; 5544-
CC 5633; 5919-6371; and 6372-9372. The sequence is also disclosed in BP-
CC 463848 in which a virus particle consg. antigens encoded by the sequence
CC is claimed, as well as expression vectors contg. the sequence. See
CC AAQ20268 for details of this specification. (Updated on 25-MAR-2003 to
CC correct PA field.)
XX
SQ Sequence 9416 BP; 1905 A; 2825 C; 2679 G; 2007 T; 0 U; 0 Other;
Query Match 87.9%; Score 1796.6; DB 2; Length 9416;
Best Local Similarity 92.5%; Pred. No. 0;
Matches 1889; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

```

```

QY 241 AACAACCCATTAACCAATTATCATGAGCATGATGACCCGACCTGGAGTGTACG 300
QY |||||
DB ACCACCCCATTAACCAATTATCATGAGCATGATGACCTGGAGTGTACGACT 5303
QY AGCACTGGGAGTGTGGGGGGGGTCTTCAGCTGCTGGCTATTTGCTTGCAACA 360
QY |||||
DB AGCACTGGGAGTGTGGGGGGGGTCTTCAGCTGCTGGCTATTTGCTTGCAACA 5363
QY 361 GGCAGCGTGTATTTGGGTAGGATCATCTTGTCCGGGCGCGGCTATTTGTCGAC 420
QY |||||
DB AGGAGCTTCTCTACAGAGATTCATAGATGGAAGTGGCGGTGCTCCTCTTAC 480
QY 421 AGGAGTCTCTCTACAGAGATTCATAGATGGAAGTGGCGGTGCTCCTCTTAC 480
QY |||||
DB AGGAGCTTCTCTACAGAGATTCATAGATGGAAGTGGCGGTGCTCCTCTTAC 5483
QY 481 ATCGAGCGAGGAATGACCTGCGAGCATTTCAAGCAAAAGCGCTGGGTGTCGAG 540
QY |||||
DB ATCGAGCGAGGAATGACCTGCGAGCATTTCAAGCAAAAGCGCTGGGTGTCGAG 5543
QY 541 ACAGCCACCAAGCAAGCGAGCGCTGCTCCGTGTGAGTCCAACTGGCGAGCCCTT 600
QY |||||
DB ACAGCCACCAAGCAAGCGAGCGCTGCTCCGTGTGAGTCCAACTGGCGAGCCCTT 5603
QY 5411 ACAGCCACCAAGCAAGCGAGCGCTGCTCCGTGTGAGTCCAACTGGCGAGCCCTT 600
QY |||||
DB ACAGCCACCAAGCAAGCGAGCGCTGCTCCGTGTGAGTCCAACTGGCGAGCCCTT 5603
QY 601 GAGACCTTCTGGGCGAAACATGTGGAATTTATCAAGCGGATACATCTTACAGAGC 660
QY |||||
DB GAGACCTTCTGGGCGAAACATGTGGAATTTATCAAGCGGATACATCTTACAGAGC 5663
QY 661 TTGTTCATCTGCTGGGAGATCCCGCATTCATCATCATGATGAGCGTTCACAGCCCTGTC 720
QY |||||
DB TTATCATCTGCTGGGAGATCCCGCATTCATCATCATGATGAGCGTTCACAGCCCTTATC 5723
QY 721 ACTAGCCGCTCACCAACCAATTAACCTCTGCTTAAATCTGGGGGATGGGTAGCC 780
QY |||||
DB AOCAGCCGCTCACCAACCAATTAACCTCTGCTTAAATCTGGGGGATGGGTAGCC 5783
QY 781 GCCCACTGCTCCCCCAGATGCTGCTTCACTTTGAGGCGCGGATGCTGTGTCG 840
QY |||||
DB GCCCACTGCTCCCCCAGATGCTGCTTCACTTTGAGGCGCGGATGCTGTGTCG 5843
QY 841 GCTGTGGCAGCATTAAGGCTTGGGAGGTTGCTTGAGCATCTTGGCGGCTATGAGCA 900
QY |||||
DB GCTGTGGCAGCATTAAGGCTTGGGAGGTTGCTTGAGCATCTTGGCGGCTATGAGCA 5903
QY 901 GAGATGCGAGCGCGCTGCTGAGCTTAAAGTCATGAGCGGAGAAATCCCTCCAGAG 960
QY |||||
DB GAGATGCGAGCGCGCTGCTGAGCTTAAAGTCATGAGCGGAGAAATCCCTCCAGAG 5963
QY 961 GACCTGTTAACTTAATCTCCCTGACATCTCTCTGCTGCTGCTGCTGCTGCTGCTG 1020
QY |||||
DB GACCTGTTAACTTAATCTCCCTGACATCTCTCTGCTGCTGCTGCTGCTGCTGCTG 6023
QY 1021 TGCGAGGATTAAGTGGTGGGAGCGTGGAGCGTGGAGGAGGAGGCTGTGAGTGAAC 1080
QY |||||
DB TGCGAGGATTAAGTGGTGGGAGCGTGGAGCGTGGAGGAGGAGGCTGTGAGTGAAC 6083
QY 6084 CGGCTGATAGGCTTGGCTGCTGGGGGTAAATATGTTTCCCAAGCACTATGTGCTGAG 6143
QY |||||
DB CGGCTGATAGGCTTGGCTGCTGGGGGTAAATATGTTTCCCAAGCACTATGTGCTGAG 6143
QY 1141 AGCGAGCGCGAGCGTGTCACTGATCTCTTCCGACCTTATCATCAACCACTGTTG 1200
QY |||||
DB AGCGAGCGCGAGCGTGTCACTGATCTCTTCCGACCTTATCATCAACCACTGTTG 6203
QY 1201 AAGAGGCTCAACATGATTAAGAGAGACTGCTCAAGCCCTGCTCGGCTCGGAGCTA 1260
QY |||||
DB AAGAGGCTCAACATGATTAAGAGAGACTGCTCAAGCCCTGCTCGGCTCGGAGCTA 6263
QY 6204 AAAAGGCTCAACATGATTAAGAGAGACTGCTCAAGCCCTGCTCGGAGCTA 6263
QY |||||
DB AAGGATGTTGGAGCTGATATGCAAGTTTGTGCTGCTTCAAGACCTGCTCCAGTCC 1320
QY 1261 AAGGATGTTGGAGCTGATATGCAAGTTTGTGCTGCTTCAAGACCTGCTCCAGTCC 1320
QY |||||
DB AAGGATGTTGGAGCTGATATGCAAGTTTGTGCTGCTTCAAGACCTGCTCCAGTCC 6323

```



```

QY 1201 AAGAGCTCCACAGATGATTAACGAGAGCTGCTCCAGCCTGCTCCGGCTGTTGAGCTA 1260
DB 6204 AAAAGGCTCCACAGATGATTAATGAAAGCTGCTCCAGCCGGTTCCTGGCTGTTGAGCTA 1263
QY 1261 AAGGATGTTTGGGATCTGGAATATGACAGTTTGGCTGACTTCAAGACCTGGCTCCAGTCC 1320
DB 6264 AAGGATGTTTGGGATCTGGAATATGACAGTTTGGCTGACTTCAAGACCTGGCTCCAGTCC 1323
QY 1321 AAGCTCTGCGCGAGTTACCGGAGTCCCTTTTCTGATGCGCAAGTGGTATCAAGAGG 1380
DB 6324 AAGCTCTGCGCGAGTTACCGGAGTCCCTTTTCTGATGCGCAAGTGGTATCAAGAGG 1383
QY 1381 GTCTGCGGAGAGAGAGCATCATGACAGACCTGCTCATGTGAGACAGATCAACCGGA 1440
DB 6384 GTCTGCGGAGAGAGAGCATCATGACAGACCTGCTCATGTGAGACAGATCAACCGGA 1443
QY 1441 CATGTCAAAAAGGTTTCCATGAGATGTTGGGCTTAAGAAGCTGTATGATCAATGTGGCAT 1500
DB 6444 CATGTCAAAAAGGTTTCCATGAGATGTTGGGCTTAAGAAGCTGTATGATCAATGTGGCAT 1503
QY 1501 GGAACATTTCCCATCAAGCATACACAGAGGCGCTGACGCGCTCCAGGCGCAAC 1560
DB 6504 GGAACATTTCCCATCAAGCATACACAGAGGCGCTGACGCGCTCCAGGCGCAAC 1563
QY 1561 TATTCAGGCGCTGTGCGGCTGCTGCTGAGAGTACGTGAGGTTACGCGGCTGAGG 1620
DB 6564 TATTCAGGCGCTGTGCGGCTGCTGCTGAGAGTACGTGAGGTTACGCGGCTGAGG 1623
QY 1621 GATTTCCACTACGTGACGACATGACCACTGACCAAGTAAATGCCCTGCGAGGTTCCA 1680
DB 6624 GATTTCCACTACGTGACGACATGACCACTGACCAAGTAAATGCCCTGCGAGGTTCCG 1683
QY 1681 GCGCCGAAATTTTCAAGAAATGAGATGGGGTGGCGGCTGACAGAGTACGCTCCGCGTGC 1740
DB 6684 GCGCCGAAATTTTCTCGAGGTGAGCGAGTGCAGAGTACGCTCCGCGTGC 1743
QY 1741 AAACCTCTCTCAAGGAGAGGTCATTCAGAGTGCAGGCTCAACCAATACCTGTTGGG 1800
DB 6744 AGGCTCTCTCAAGGAGAGGTCATTCAGAGTGCAGGCTCAACCAATACCTGTTGGG 1803
QY 1801 TCGAGCTTCCATGAGAGCCCAACCGGATGATGACAGTGTCTCACTTCATGCTCACGAC 1860
DB 6804 TCGAGCTTCCATGAGAGCCCAACCGGATGATGACAGTGTCTCACTTCATGCTCACGAC 1863
QY 1861 CCTCCCATCAACGAGAGAGCGGTAGCCGAGGCTGCGAGGCGGCTCCCGCTCC 1920
DB 6864 CCTCCCATCAACGAGAGAGCGGTAGCCGAGGCTGCGAGGCGGCTCCCGCTCC 1923
QY 1921 TTGGCAGCTCTTCAAGTACAGTGTGTCGCGCTTCTCGAAGGCGACATACATTACC 1980
DB 6924 TTGGCAGCTCTTCAAGTACAGTGTGTCGCGCTTCTCGAAGGCGACATACATTACC 1983
QY 1981 CAAAATGACTTCCAGAGCGCTGACCTCATGAGGCGCAACTTCCTGCGGCGATGAGATG 2040
DB 6984 CACCATGCTCTCCGAGCGCTGACCTCATGAGGCGCAACTTCCTGCGGCGAGGATG 2043
QY 2041 GGC 2043
DB 7044 GGC 7046

```

RESULT 11
AAD25321
AAD25321 standard: cDNA, 7987 BP.

AC AAD25321,
DT 12-MAR-2002 (first entry)
XX Hepatitis C virus (HCV) repliBartman/delta2U's cDNA.
DE Hepatitis C virus (HCV) repliBartman/delta2U's cDNA.
XX Hepatitis C virus; HCV; transfection; infection; virus neutralisation;
KW gene therapy; vaccine; immunoprotection; hepatotropic; virucide; liver;

```

KW 88.
XX Hepatitis C virus.
OS WO200189364-A2.
PN 29-NOV-2001.
XX 23-MAY-2001; 2001WO-US016822.
XX 23-MAY-2000; 2000US-00576989.
XX (UNITW) UNITV WASHINGTON.
XX Rice CM, Blight KJ;
XX WPI, 2002-06755/09.
DR Hepatitis C virus variants having greater transfection efficiency and
PT ability to survive subpassage, useful as a vaccine for immunizing primate
PT to the virus, comprise non-naturally occurring viral sequences.
PS Claim 44; Page 66-69; 174pp; English.
XX The invention relates to Hepatitis C virus (HCV) variants which include
XX polynucleotides comprising non-naturally occurring HCV sequence and HCV
XX variants that have a transfection efficiency and ability to survive
XX subpassage greater than HCV that have wild-type polypeptide coding
XX regions. The polynucleotides of the invention are useful for identifying
XX a cell line that is permissive for infection with HCV and detecting
XX replication of HCV in cells of the cell line. They are also useful for
XX testing a compound for anti-viral properties and for inhibiting HCV
XX infection. They are also useful for the generation of defined HCV virus
XX stocks to develop in vitro and in vivo assays for virus neutralisation,
XX attachment, penetration and entry, structure/function studies on HCV
XX proteins and RNA elements and identification of new antiviral targets, a
XX systematic survey of cell culture systems and conditions to identify
XX those that support wild-type and variant HCV RNA replication and particle
XX release, production of adaptive HCV variants capable of more efficiency
XX replication in cell culture, production of HCV variants with altered
XX tissue or species tropism, establishment of alternative animal models for
XX inhibitor evaluation including those supporting HCV variant replication,
XX development of cell-free HCV replication assays, production of
XX immunogenic HCV particles for vaccination, engineering of attenuated HCV
XX derivatives as possible vaccine candidates, engineering of attenuated or
XX defective HCV derivatives for expression of heterologous gene products
XX for gene therapy and vaccine applications and for utilisation of the HCV
XX glycoproteins for targeted delivery of therapeutic agents to the liver
XX or other cell types with appropriate receptors. Vaccine comprising these
XX sequences is useful for inducing immunoprotection to HCV in a primate.
XX The present sequence is Hepatitis C virus (HCV) repliBartman/delta2U's
XX cDNA
SQ Sequence 7987 BP; 1647 A; 2368 C; 2243 G; 1729 T; 0 U; 0 Other;
XX
XX Query Match 87.5%; Score 1787; DB 6; Length 7987;
XX Best Local Similarity 92.2%; Pred. No. 0; Mismatches 0; Gaps 0;
XX Matches 1893; Conservative 0; Indels 0;
QY 1 TGGAGGCGCTTTCACAGGCTTCCACGAGTGAATGCGCACTTCTGCTCCCAACAAAG 60
DB 3397 TGGAGAGAGGTGTTTACAGGCTTCCACGAGTGAATGCGCACTTCTGCTCCCAACAAAG 3456
QY 61 CAGGAGAGAGACATTTCCCTTCACTGTTGGGCTTCCAGGCTACTGTGTGGGCTTGGGCT 120
DB 3457 CAGGAGAGAGACATTTCCCTTCACTGTTGGGCTTCCAGGCTACTGTGTGGGCTTGGGCT 3516
QY 121 CAGGCGCCACCTCCATCAATGAGATGAGAGTGTCTCATGAGGCTAAAGCTACT 180
DB 3517 CAGGCTCCACCTTCACTGTTGGGCTTCCAGGAGTGTCTCATGAGGCTAAAGCTACTAG 3576
QY 181 CTGCGCGGCGCAACCTTGTGTATGAGTGTGGAGCGCTTCAACAGAGGTCACTC 240

```


XX 03-APR-1999; 99DE-01015178.
 PR
 XX (UNWA-) UNIV MAINZ GUTENBERG JOHANNES.
 PA
 XX Bartenschlager R;
 PI
 XX WPI, 2000-629140/61.
 DR
 XX Cell culture system for hepatitis C virus, useful e.g. in screening for
 PT therapeutic agents, comprises human hepatoma cells containing a viral RNA
 PT construct that includes a selectable gene.
 XX
 PS Claim 8; Page 37-43; 58pp; German.
 XX
 CC This invention describes a novel Hepatitis C virus (HCV) cell culture
 CC system comprising human hepatoma cells that contain an integrated HCV-RNA
 CC construct (I). (1) contains the HCV-specific RNA segments 5'-NTR (non-
 CC translated region), NS (non-structural)3, NS4A, NS4B, NS5A, NS5B and 3'-
 CC NTR, and a selectable (marker) gene (II). The cell cultures, and/or (I),
 CC are used to prepare, evaluate and/or test therapeutic and/or diagnostic
 CC agents for HCV infections, and to prepare vaccines against HCV infection
 CC (particularly preparation of attenuated HCV). The can also be used for
 CC preparation of a liver-specific delivery system for gene therapy, and to
 CC identify cells permissive for HCV replication. Virus RNA replicates
 CC autonomously and with high efficiency in this cellular system, so that
 CC variations in replication rates can be measured (for screening antiviral
 CC agents) quantitatively or qualitatively, using standard laboratory
 CC equipment. Efficient replication of HCV RNA is only achieved when the
 CC specified RNA segments are present and when the transfected cells are
 CC maintained under permanent selection pressure
 CC
 SO Sequence 7989 BP; 1647 A; 2368 C; 2243 G; 1731 T; 0 U; 0 Other;
 Query Match 87.5%; Score 1787; DB 3; Length 7989;
 Beest Local Similarity 92.2%; Pred. No. 0;
 Matches 1883; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

QY 541 ACAGCCACCAAGCAAGCGAGGCGCTGCTCCCGTGTGAGTCCAGTGGCAGCCCTT 600
 DB 3937 ACAGCCACCAAGCAAGCGAGGCGCTGCTCCCGTGTGAGTCCAGTGGCAGCCCTC 3996
 QY 601 GAGACCTTGTGGGGGAAACATGTGAACTTCATCAGCGGATACAGTCTTAGCAGGC 660
 DB 3997 GAGACCTTGTGGGGGAAACATGTGAACTTCATCAGCGGATACAGTCTTAGCAGGC 4056
 QY 661 TTGTCCACTGCGCGGAAATCCCGGATGTGATACGATGCGGTTACAGCCTGTGC 720
 DB 4057 TTGTCCACTGCGCGGAAATCCCGGATGTGATACGATGCGGTTACAGCCTGTATC 4116
 QY 721 ACTAGCCGCTCACCAACCAATCTACCTCTGCTTAAATCTCTGGGGGATGAGTAC 780
 DB 4117 ACCAGCCGCTCACCAACCAATCTCTGCTTAAATCTCTGGGGGATGAGTAC 4176
 QY 781 GCCCACTGCTCCCGGAGTGTCTTCAAGTGTGAGCGCGCATTTGCTGTGC 840
 DB 4177 GCCCACTGCTCCCGGAGTGTCTTCAAGTGTGAGCGCGCATTTGCTGTGAGCG 4236
 QY 841 GCTTTGGCAGATAGGCTTGGGAAAGTGTGAGCATCTTGGCGGCTATGAGCA 900
 DB 4237 GCTTTGGCAGATAGGCTTGGGAAAGTGTGAGCATCTTGGCGGCTATGAGCA 4296
 QY 901 GGAGTGGCAGGCGCGCTGCTGAGCTTTAAGTATGAGCGGGAATCCCTCCAGCAG 960
 DB 4297 GGAGTGGCAGGCGCGCTGCTGAGCTTTAAGTATGAGCGGGAATCCCTCCAGCAG 4356
 QY 961 GACCTGTAACTTATCTCCCTGCACTCTCTCTGTGTCCTGTGCTGTGCTGTG 1020
 DB 4357 GACCTGTAACTTATCTCCCTGCACTCTCTCTGTGTCCTGTGCTGTGCTGTG 4416
 QY 1021 TGGCAGGATATCTGCTGCGACGTGGGTCCAGGGGAGGGGCTGTGAGTGAAC 1080
 DB 4417 TGGCAGGATATCTGCTGCGACGTGGGTCCAGGGGAGGGGCTGTGAGTGAAC 4476
 QY 1081 CGGCTGATAGGCTTGGCGCTGCGGGGTTAACATGTTTCCCGCAGCATATGCAAG 1140
 DB 4477 CGGCTGATAGGCTTGGCGCTGCGGGGTTAACATGTTTCCCGCAGCATATGCAAG 4536
 QY 1141 AGCGAGCGCGCAGCAGTGTCTACTAGATCTCTCCGACTTATCATACCACTGTG 1200
 DB 4537 AGCGAGCGCGCAGCAGTGTCTACTAGATCTCTCCGACTTATCATACCACTGTG 4596
 QY 1201 AAGAGCTCACAGTGAATTAAGAGATGTCTCAAGCCTGCTCGGCTGTGCTGA 1260
 DB 4597 AAGAGCTCACAGTGAATTAAGAGATGTCTCAAGCCTGCTCGGCTGTGCTGA 4656
 QY 1261 AGGATGTTTGGGACTGATATGCAAGTGTGCTCAAGACCTGCTCGAGTCC 1320
 DB 4657 AGGATGTTTGGGACTGATATGCAAGTGTGCTCAAGACCTGCTCGAGTCC 4716
 QY 1321 AAGCTCTGCGCGATTAACCGGAGTCCCTTTTCTCATGCAACGTGGTACAAGGG 1380
 DB 4717 AAGCTCTGCGCGATTAACCGGAGTCCCTTTTCTCATGCAACGTGGTACAAGGG 4776
 QY 1381 GTTGGCGGGAGAGCGGATCATGAGACCACTGCTCATGTGAGCAAGATCACCGGA 1440
 DB 4777 GTTGGCGGGAGAGCGGATCATGAGACCACTGCTCATGTGAGCAAGATCACCGGA 4836
 QY 1441 CATGTCAAAAAGGTTTCATGAGATGTTGGGCTTAAGACCTGTATGTAATGTGGAT 1500
 DB 4837 CATGTCAAAAAGGTTTCATGAGATGTTGGGCTTAAGACCTGTATGTAATGTGGAT 4896
 QY 1501 GGAACATTCCTCATCAACGATACACAGGGGCTCTGACGCTCCAGGCGCAAC 1560
 DB 4897 GGAACATTCCTCATCAACGATACACAGGGGCTCTGACGCTCCAGGCGCAAC 4956
 QY 1561 TATTCAGGCGCTGTGCGGGTGTGCTGAGAGTACGTGAGTTACGGGGTGGGG 1620
 DB 4957 TATTCAGGCGCTGTGCGGGTGTGCTGAGAGTACGTGAGTTACGGGGTGGGG 5016

QY 1621 GATTTCACATAGTACGAGCATGACCACTGACCAACGTAAATGCCCGTCCAGGTTCCA 1680
DB 5017 GATTTCACATAGTACGAGCATGACCACTGACCAACGTAAATGCCCGTCCAGGTTCCA 5076
QY 1681 GCCCGGCAATCTTACAGAAAGTGGATGGGCTGCGGCTGCAAGGATGACCTCGGCGTGC 1740
DB 5077 GCCCGGCAATCTTACAGAAAGTGGATGGGCTGCGGCTGCAAGGATGACCTCGGCGTGC 5136
QY 1741 AAAACCTCTCTTACGAGGAGAGGTCACTTCCAGTCCGAGCTGACCAATACCTGGTTGGG 1800
DB 5137 AAAACCTCTCTTACGAGGAGAGGTCACTTCCAGTCCGAGCTGACCAATACCTGGTTGGG 5196
QY 1801 TCGAGAGCTCCATGCGAGAGCCCAACCGGATGTAGACAGTCTCACTTCCATGCTCACGAC 1860
DB 5197 TCGAGAGCTCCATGCGAGAGCCCAACCGGATGTAGACAGTCTCACTTCCATGCTCACGAC 5256
QY 1861 CCTCCCATCATGACAGACAGAGCGGCTAAAGCGAGGCTGAGCGAGGAGGCTGCCCCCTCC 1920
DB 5257 CCTCCCATCATGACAGAGCGGCTAAAGCGAGGCTGAGCGAGGAGGCTGCCCCCTCC 5316
QY 1921 TTGGCCAGCTCTTACGAGTACGAGGCTGAGCGGCTTCTCTCGAAGGCGACATATACCTAAC 1980
DB 5317 TTGGCCAGCTCTTACGAGTACGAGGCTGAGCGGCTTCTCTCGAAGGCGACATATACCTAAC 5376
QY 1981 CAAATGACTTCCGAGAGCGCTGACCTCATTCGAGGCGCAACCTCTGTGGCGGATGAGATG 2040
DB 5377 CGTCAATGACTTCCGAGAGCGCTGACCTCATTCGAGGCGCAACCTCTGTGGCGGATGAGATG 5436
QY 2041 GGC 2043
DB 5437 GGC 5439

RESULT 13

AAD25322
ID AAD25322 standard; cDNA; 7989 BP.

XX AAD25322;

XX 12-MAR-2002 (first entry)

XX Hepatitis C virus (HCV) repliBartman/Availi cDNA.

XX Hepatitis C virus; HCV; transfection; infection; virus neutralization;

XX gene therapy; vaccine; immunoprotection; hepatotropic; virucide; liver;

XX Hepatitis C virus.

XX Key Location/Qualifiers

XX CDS 1801..7758

XX misc_feature 7766

XX /tag= b

XX /note= "Nucleotide creating Availi site"

XX MO200189364-A2.

XX 29-NOV-2001.

XX 23-MAY-2001; 2001WO-US016822.

XX 23-MAY-2000; 2000US-00576989.

XX (UNIM) UNIV WASHINGTON.

XX Rice CM, Blight KJ;

XX MPI; 2002-066755/09.

XX P-PSDB; AAE15717.

XX Hepatitis C virus variants having greater transfection efficiency and

PT ability to survive subpassage, useful as a vaccine for immunizing primate
PT to the virus, comprise non-naturally occurring viral sequences.
PS Claim 44; Page 69-71; 1749p; English.

CC The invention relates to Hepatitis C virus (HCV) variants which include
CC polynucleotides comprising non-naturally occurring HCV sequence and HCV
CC variants that have a transfection efficiency and ability to survive
CC subpassage greater than HCV that have wild-type polypeptide coding
CC regions. The polynucleotides of the invention are useful for identifying
CC a cell line that is permissive for infection with HCV and detecting
CC replication of HCV in cells of the cell line. They are also useful for
CC testing a compound for anti-viral properties and for inhibiting HCV
CC infection. They are also useful for the generation of defined HCV virus
CC stocks to develop in vitro and in vivo assays for virus neutralisation,
CC attachment, penetration and entry, structure/function studies on HCV
CC proteins and RNA elements and identification of new antiviral targets, a
CC systematic survey of cell culture systems and conditions to identify
CC those that support wild-type and variant HCV RNA replication and particle
CC release, production of adaptive HCV variants capable of more efficiency
CC replication in cell culture, production of HCV variants with altered
CC tissue or species tropism, establishment of alternative animal models for
CC inhibitor evaluation including those supporting HCV variant replication,
CC development of cell-free HCV replication assays, production of
CC immunogenic HCV particles for vaccination, engineering of attenuated HCV
CC derivatives as possible vaccine candidates, engineering of attenuated or
CC defective HCV derivatives for expression of heterologous gene products
CC for gene therapy and vaccine applications and for utilisation of the HCV
CC glycoproteins for targeted delivery of therapeutic agents to the liver
CC or other cell types with appropriate receptors. Vaccine comprising these
CC sequences is useful for inducing immunoprotection to HCV in a primate.
CC The present sequence is Hepatitis C virus (HCV) repliBartman/Availi cDNA

Sequence 7989 BP; 1647 A; 2369 C; 2242 G; 1731 T; 0 U; 0 Other;

Query Match 87.5%; Score 1787; DB 6; Length 7989;

Best Local Similarity 92.2%; Pred. No. 0; Mismatches 160; Indels 0; Gaps 0;

Matches 1883; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

QY 1 TGGAGAGGAGTCTTACAGAGCTGACCCAGCTGAGTGGCCACTTCTCCCAACAG 60
DB 3397 TGGAGAGGAGTCTTACAGAGCTGACCCAGCTGAGTGGCCACTTCTCCCAACAG 3456
QY 61 CAGGAGAGAGCAACATCTCCCTACCTGAGCTGACAGGCTGAGTGGCTGAGGCTC 120
DB 3457 CAGGAGAGAGCAACATCTCCCTACCTGAGCTGAGTGGCTGAGTGGCTGAGGCTC 3516
QY 121 CAGGAGAGAGCAACATCTCCCTACCTGAGCTGAGTGGCTGAGTGGCTGAGGCTC 180
DB 3517 CAGGAGAGAGCAACATCTCCCTACCTGAGCTGAGTGGCTGAGTGGCTGAGGCTC 3576
QY 181 CTGCGGAGAGCAACATCTCCCTGATGAGCTGAGGAGCGGCTCAAAAGAGGCTCACCTC 240
DB 3577 CTGCGGAGAGCAACATCTCCCTGATGAGCTGAGGAGCGGCTCAAAAGAGGCTCACCTC 3636
QY 241 ACACACCCCAATACCAATATCATCATGAGTGCATGTCAGGCGACCTGAGGCTGCTCAG 300
DB 3637 ACACACCCCAATACCAATATCATCATGAGTGCATGTCAGGCGACCTGAGGCTGCTCAG 3696
QY 301 AGCAGCTGGGCTGAGTGGGAGGAGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTG 360
DB 3697 AGCAGCTGGGCTGAGTGGGAGGAGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTG 3756
QY 361 GGCAGGCTGATCATTTGAGTGGAGTCACTTTGTCGAGGAGGAGGAGGAGGAGGAGGAG 420
DB 3757 GGCAGGCTGATCATTTGAGTGGAGTCACTTTGTCGAGGAGGAGGAGGAGGAGGAGG 3816
QY 421 AGGAGAGTCTTACCAAGAGTTCAGTGAAGAGTGCAGGCTGAGGCTGAGGCTGAGGCT 480
DB 3817 AGGAGAGTCTTACCAAGAGTTCAGTGAAGAGTGCAGGCTGAGGCTGAGGCTGAGGCT 3876
QY 481 ATCGAGAGGAGTACAGTGGCGAGAGTTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 540

```

Db 3877 ATCGAACAAGGATGAGCTCCGCAACATTCMAACAGAGGCAATCGGTTGCTGCA 3936
Qy 541 AAGGCAACCAAGCAAGCGGAGCGCTGCTCCGTTGGAGTCCAAAGTGGGAGCCCTT 600
Db 3937 ACAGCCACCAACCAAGCGGAGCTGCTCCGTTGGAGTCCAAAGTGGGAGCCCTT 3996
Qy 601 GAGACCTTCTGGGCGCAACACATGTGGAATTATCAGCGGGAATACATTAATTAGCAGC 660
Db 3997 GAAGCTTCTGGGCGCAAGATATGTGGAATTCATCAGCGGGAATACATTAATTAGCAGC 4056
Qy 661 TTGTCACTCTGCTGGGGAATCCCGGATTCATCATGTGCGCTTCACAGCTCTGTTC 720
Db 4057 TTGTCACTCTGCTGGGGAATCCCGGATTCATCATGTGCGCTTCACAGCTCTGTTC 4116
Qy 721 ACTAGCCGCTCACCACCAATCTACCCCTGCTTAATCATCTGGGGGATGGGATGCG 780
Db 4117 ACCACCCGCTCACCACCAATCTACCCCTGCTTAATCATCTGGGGGATGGGATGCG 4176
Qy 781 GCCCACTCGTCCCCCAGTCTGCTGAGCTTTCTGAGGCGCCGCAATTCGCTGTCG 840
Db 4177 GCCCACTCGTCCCCCAGTCTGCTGAGCTTTCTGAGGCGCCGCAATTCGCTGTCG 4236
Qy 841 GCTGTGGCAGCATAGGCTTTGGAAAGTGTGTGTGACATCTTGCGGGCTATGAGCA 900
Db 4237 GCTGTGGCAGCATAGGCTTTGGAAAGTGTGTGTGATATTTGGCAAGTTATGAGCA 4296
Qy 901 GGAAGGCGAGGCGCTGTGGCTTTAGGTATGAGGCGGGAATGCCCTCCACCGAG 960
Db 4297 GGAAGGCGAGGCGCTGTGGCTTTAGGTATGAGGCGGGAATGCCCTCCACCGAG 960
Qy 961 GACCTGTAACTTCTCCCTGCAATCTCTCTCTGTGTGCTGCTGCTGCTGCTGCTG 1020
Db 4357 GACCTGTAACTTCTCCCTGCAATCTCTCTCTGTGTGCTGCTGCTGCTGCTGCTG 4416
Qy 1021 TCGCAGCGCATACTGCTGCGCACTGTGGTCAAGGCGGAGGCGCTGTGAGTGAAC 1080
Db 4417 TCGCAGCGCATACTGCTGCGCACTGTGGTCAAGGCGGAGGCGCTGTGAGTGAAC 4476
Qy 1081 CGGCTGATAGCGCTTCCGCGGGGATACATGTTTCCCGCAAGCATATGTCGAGAG 1140
Db 4477 CGGCTGATAGCGCTTCCGCGGGGATACATGTTTCCCGCAAGCATATGTCGAGAG 4536
Qy 1141 AGCGAGCGCGAGCATGTCTCATCTGATCTCTCCGACTTACTATCACTCACTGTTG 1200
Db 4537 AGCGAGCGCGAGCATGTCTCATCTGATCTCTCCGACTTACTATCACTCACTGTTG 4596
Qy 1201 AAGAGGCTCCACAGTGAATTAACAGAGACTGCTCCAGCGCTCGGCTCGTGGCTA 1260
Db 4597 AAGAGGCTCCACAGTGAATTAACAGAGACTGCTCCAGCGCTCGGCTCGTGGCTA 4656
Qy 1261 AAGGATGTTTGGGATCTGATATGCAAGTGTGGCTGACTTCAAGACCTGGCTCCAGTCC 1320
Db 4657 AAGGATGTTTGGGATCTGATATGCAAGTGTGGCTGACTTCAAGACCTGGCTCCAGTCC 4716
Qy 1321 AAGCTCTGCGCGATTAACCGGAGTCCCTTTTCTCATGCAACGTGGTACAAGGGG 1380
Db 4717 AAGCTCTGCGCGATTAACCGGAGTCCCTTTTCTCATGCAACGTGGTACAAGGGG 4776
Qy 1381 GTCTGGCGGGAGAGCGCATCAATGACACCTGCTCATGTGAGCAGATCAACCGGA 1440
Db 4777 GTCTGGCGGGAGAGCGCATCAATGACACCTGCTCATGTGAGCAGATCAACCGGA 4836
Qy 1441 CATGTCAAAAACGGTTCATGAGATCGTTGGGCTTAAGACTGTATTAACATGTGGCAT 1500
Db 4837 CATGTCAAAAACGGTTCATGAGATCGTTGGGCTTAAGACTGTATTAACATGTGGCAT 4896
Qy 1501 GGAACATTCCTCCATCAACGATACACAGGCGCCCTGACGCGCTCCCGAGCGCAAC 1560
Db 4897 GGAACATTCCTCCATCAACGATACACAGGCGCCCTGACGCGCTCCCGAGCGCAAC 4956
Qy 1561 TATTCAGGCGCTGTGGCGGTGTGCTGTAAGAGATGATGAGAGTTACGCGGGTGGG 1620
Db 4957 TATTCAGGCGCTGTGGCGGTGTGCTGTAAGAGATGATGAGAGTTACGCGGGTGGG 5016

```

```

Qy 1621 GATTTCACTACGTAGCAGCATGACCACTGACAAAGTAAATGCCCGTCCAGTTCCA 1680
Db 5017 GATTTCACTACGTAGCAGCATGACCACTGACAAAGTAAATGCCCGTCCAGTTCCA 5076
Qy 1681 GCCCGGAATTTCTTCAACAGAGTGAATGGGCTGCGCTGACAGGTAAGCTCCGCGTCC 1740
Db 5077 GCCCGGAATTTCTTCAACAGAGTGAATGGGCTGCGCTGACAGGTAAGCTCCGCGTCC 5136
Qy 1741 AAACCTTCCTACGGAAGAGTCAATTCAGTCCGCTGCTGCTGCTGCTGCTGCTGCTG 1800
Db 5137 AAACCTTCCTACGGAAGAGTCAATTCAGTCCGCTGCTGCTGCTGCTGCTGCTGCTG 5196
Qy 1801 TCGAGCTCCCATGAGCGAGCCGCAACCGAGTATGAGAGTCACTTCCATGCTCAACGAC 1860
Db 5197 TCGAGCTCCCATGAGCGAGCCGCAACCGAGTATGAGAGTCACTTCCATGCTCAACGAC 5256
Qy 1861 CCTCCCATGAGCAGCAGAGCGCTTACGCGAGGCTGGCCAGGAGGCTTCCCTCCCTCC 1920
Db 5257 CCTCCCATGAGCAGCAGAGCGCTTACGCGAGGCTGGCCAGGAGGCTTCCCTCCCTCC 5316
Qy 1921 TTGGCAGCTCTTACGCTAGCCAGTGTGTGCTGCTTCTCTGCAAGGCAATACATTAAC 1980
Db 5317 TTGGCAGCTCTTACGCTAGCCAGTGTGTGCTGCTTCTCTGCAAGGCAATACATTAAC 5376
Qy 1981 CAAATGACTTCCAGAGCGCTGACCTCATCGAGGCAACCTCTGCTGGCGGATGAGATG 2040
Db 5377 CCAATGACTTCCAGAGCGCTGACCTCATCGAGGCAACCTCTGCTGGCGGATGAGATG 5436
Qy 2041 GGC 2043
Db 5437 GGC 5439

```

```

RESULT 14
ADJ57845
ID ADJ57845 strand; DNA; 7989 BP.
XX
AC ADJ57845;
XX
DT 06-MAY-2004 (first entry)
XX
DE HCV replicon encoding sequence.
XX
HE hepatitis C virus; HCV; Antinflammatory; Hepatotropic; Virucide; de;
XX HCV replicon.
XX
OS unidentified.
XX
FT Key location/Qualifiers
FT CDS 1801..7759
FT /tag= a
FT /product= "HCV replicon"
XX
PN MO200401511-A2.
XX
PD 19-FEB-2004.
XX
PF 12-AUG-2003; 2003WO-US025260.
XX
PR 12-AUG-2002; 2002US-040261P.
XX
PA (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
PI Gao M, Lemm JA, O'Boyle DR, Nower P, Rigat K, Sun J;
XX
DR MPI; 2004-180685/17.
XX
P-PSDB; ADJ57846.
XX
PT Use of hepatitis C virus assays or reporter assays, e.g. identifying a
PT compound that inhibits hepatitis C virus RNA replication or identifying a
PT compound that modulates the activity of a gene of interest.
XX

```

PS Claim 3, SEQ ID NO 1; 45bp; English.
XX The present invention relates to the use of hepatitis C virus (HCV)
CC assays for identifying a compound that inhibits HCV RNA replication and
CC reporter assays for identifying a compound that modulates the activity of
CC a gene of interest. The assays are useful for identifying a compound that
CC inhibits HCV RNA replication or for identifying a compound that modulates
CC the activity of a gene of interest. The HCV assay is useful for high
CC throughput screening that quantifies both the amount of HCV RNA
CC replication inhibitory activity associated with a test compound and the
CC amount of cytotoxicity associated with the test compound. The compound is
CC useful for treating hepatitis C infection. Assays of the invention have
CC distinct advantages when compared to qRT-PCR or other methods in that
CC assays of the invention may take place *in situ* in a detergent based crude
CC cell lysate, which requires no further preparation prior to performing
CC the assays. The assays do not also involve numerous manipulations to add
CC or subtract reagents after addition of test compounds and are desirably
CC based on a viral protein which is required by the HCV replicon for
CC replication. The present sequence represents a HCV replicon encoding
CC sequence used in the assay of the invention.
XX

Sequence 7989 BP; 1647 A; 2368 C; 2243 G; 1731 T; 0 U; 0 Other;

Query Match 87.5%; Score 1787; DB 12; Length 7989;

Best Local Similarity 92.2%; Pred.No. 0;

Matches 1883; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

```
OY 1 TGGGAGGGGCGTCTTACAGAGGCTCAACCCGATGAGTGCACCTTCCGCAACAAG 60
DB 3397 TGGGAGAGCGCTTTACAGGCTTCAACCCGATGAGTGCACCTTCTTCCAGACTAG 3456
OY 61 CAGGAGAGAGCAACTTCCCTACCTGATGAGCGGTAACAGGCTATCTGTGCGCTAAGGCC 120
DB 3457 CAGGAGAGAGCAACTTCCCTACCTGATGAGCGGTAACAGGCTATCTGTGCGCGCTAAGGCC 3516
OY 121 CAGGCGCCACCTTCCATCATGAGTCAAAATGTGGAAGTGTCTCATACGGCTAAAGCTTACT 180
DB 3517 CAGGCTCCACCTTCCATCATGAGTCAAAATGTGGAAGTGTCTCATACGGCTAAAGCTTACT 3576
OY 181 CTGGCGGGGCGCAACCCCTTGTGTATAGGCTGGGAGCGGCTCAAAAGAGGCTCAACCTTC 240
DB 3577 CTGGCGGGGCGCAACCCCTTGTGTATAGGCTGGGAGCGGCTCAAAAGAGGCTTACTTAC 3636
OY 241 ACACACCCCATTAACCAATTCATCATGAGTGCATGTCTACGACCTGAGAGGTCTGTACG 300
DB 3637 ACACACCCCATTAACCAATTCATCATGAGTGCATGTCTGTGACCTGAGAGGTCTGTACG 3696
OY 301 AGCACCTGGGTCTGTGTGGGGGGGCTCTTGCAGCTTGTGGTGGTATGTCTTGACACAA 360
DB 3697 AGCACCTGGGTCTGTGTGGGGGGGCTCTTGCAGCTTGTGGTGGTATGTCTTGACACAA 3756
OY 361 GGCAGGCTGTATGTGTGGGTAGGATCATCTTGTCCGGGCGGCGCTATTGTTCCCGAC 420
DB 3757 GGCAGGCTGTATGTGTGGGTAGGATCATCTTGTCCGGGCGGCGCTATTGTTCCCGAC 480
OY 421 AGGGAAGTCTCTACAGAGATTGCATGAGATGGAAGTGCAGCGCTGCACCTCCCTTAC 480
DB 3811 AGGGAAGTCTCTTACAGAGATTGCATGAGATGGAAGTGCAGCGCTGCACCTCCCTTAC 3876
OY 481 ATCGAGACAGGAATGAGCTCGCCGAGCTTCAAGCAAAAAGCGCTGGGTTTCTGCGAC 540
DB 3877 ATCGAAGAGGAATGAGCTCGCCGAGCAATTCAACAGAGGCAATCGGGTTTCTGCGAA 3936
OY 541 ACAGCCACCAAGCAAGCGAGGCGCTGCTCCCGTGTGAGGTCCAAAGTGGCGAGCGCTT 600
DB 3937 ACAGCCACCAAGCAAGCGAGGCGCTGCTCCCGTGTGAGGTCCAAAGTGGCGAGCGCTT 3996
OY 601 GAGACCTTCTGGGCGGAACAACATGTGGAATTTCATCAGCGGGAATACATTAATTAGCAGCG 660
DB 3997 GAGACCTTCTGGGCGGAACAACATGTGGAATTTCATCAGCGGGAATACATTAATTAGCAGCG 4056
OY 661 TTGTCCACTCTGTGCTGGGAATCCCGGATTCATCACTGATGGCGTTCAACGCTCTGTTC 720
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
DB 4057 TTGTCCACTCTGTGCTGGGAACCCCGGATACATCATGATGAGGATTCACAGCCTCTATC 4116
OY 721 ACTAGCCGCTCACCACCACCAATCTACCTCTCTCTTAAATCTCTGGGGGATGGGTAGCC 780
DB 4117 ACCAGCCGCTCACCACCACCAATCTCTCTCTTAAATCTCTGGGGGATGGGTAGCC 4176
OY 781 GCCCACTCGTCCCCCGAGTGTCTTACCTTTCGAGGGCGCGGATGTGTGTGCG 840
DB 4177 GCCCACTCGTCCCCCGAGTGTCTTACCTTTCGAGGGCGCGGATGTGTGTGCG 4236
OY 841 GCTGTGGCAGCATAGGCTTGGGAAGGTGCTTGTGACATCTTGGCGGGCTATAGACA 900
DB 4237 GCTGTGGCAGCATAGGCTTGGGAAGGTGCTTGTGATATTTTGGCAGGTATAGACA 4296
OY 901 GAGTGTGAGCGCGCTCTGTGCGCTTTAAGTCAATAGCGGCAATATCCCTCCACCGAG 960
DB 4297 GGGGTGTGAGCGCGCTCTGTGCGCTTTAAGTCAATAGCGGCAATATCCCTCCACCGAG 4356
OY 961 GACCTGTCTTAACTTATCTCTGCAATCTCTCTCTGCTGTGCGGTGTGGGGTGTG 1020
DB 4357 GACCTGTCTTAACTTATCTCTGCAATCTCTCTCTGCTGTGCGGTGTGGGGTGTG 4416
OY 1021 TGGCAGCGATATAGCGCTGCGGACGTTGGGTCCAGGGAGGGGCTGTGACATGATGAAC 1080
DB 4417 TGGCAGCGATATAGCGCTGCGGACGTTGGGTCCAGGGAGGGGCTGTGACATGATGAAC 4476
OY 4477 CGGCTGATAGCGTTCGCTCGGGGTAAACAGCTCTCCCAAGCACTATGTGCTGAG 4536
DB 1141 AGCGACCGCGAGCAGCTGTCACTCAATCTCTCCGACCTTATCAATCAACCACTGTG 1200
OY 4537 AGCGACCGCTGAGCAGCTGTCACTCAATCTCTCTTATGATCACTCACTGCTGCTG 4596
DB 1201 AAGAGCTTCAACGATGATTAACGAGGACTGTCTCAAGCCCTGCTCGGCTCTGTGCTA 1260
OY 4597 AAGAGCTTCAACGATGATTAACGAGGACTGTCTCAAGCCCTGCTCGGCTCTGTGCTA 4656
DB 1261 AAGGATGTTTGGGACTGATATGCAAGTGTGTGCTGACTTCAAGACTGTGCTCAAGTTC 1320
OY 4657 AAGGATGTTTGGGATGATATGCAAGTGTGTGCTGACTTCAAGACTGTGCTCAAGTTC 4716
DB 1321 AAGCTCTGCGCGCATTAACGGGAGTCCCCCTTTTCTCATGCAACGTTGGGTAAAGGG 1380
OY 4717 AAGCTCTGCGCGCATTAACGGGAGTCCCCCTTTTCTCATGCAACGTTGGGTAAAGGG 4776
DB 4777 GTTGGCGGGGAGAGGATCATGAGTCAACGCTGATGTGAGGACACAGATCACCGGA 4836
OY 1381 GTTGGCGGGGAGAGGATCATGAGTCAACGCTGATGTGAGGACACAGATCACCGGA 1440
DB 4777 GTTGGCGGGGAGAGGATCATGAGTCAACGCTGATGTGAGGACACAGATCACCGGA 4836
OY 1441 CATGTCAAAAACGTTTCATGAGATGTTGGGCTTAAAGCTGTATGATATGTGGCAT 1500
DB 4837 CATGTCAAAAACGTTTCATGAGATGTTGGGCTTAAAGCTGTATGATATGTGGCAT 4896
OY 1501 GGAATATTCCTCATTAAGCATACACACGAGGCCCCCTGACAGCCCTCCCGAGCAAC 1560
DB 4897 GGAATATTCCTCATTAAGCATACACACGAGGCCCCCTGACAGCCCTCCCGAGCAAT 4956
OY 1561 TATTTCAGGGGCTGTGGCGGGTGTCTGTGAGAGTACGTGAGGTTACGGGGTGTGGG 1620
DB 4957 TATTTCAGGGGCTGTGGCGGGTGTCTGTGAGAGTACGTGAGGTTACGGGGTGTGGG 5016
OY 1621 GATTTTCACTAGTACGAGCATGACATGACAACTGTAATGCGTGTGCGAGGTTTCA 1680
DB 5017 GATTTTCACTAGTACGAGGATGACATGACAACTGTAATGCGTGTGCGAGGTTTCA 5076
OY 1681 GCCCGCAATTTCTTACAGAGTGAATGGGGTGTGCAAGATTAAGCTTCCGCGTGC 1740
DB 5077 GCCCGCAATTTCTTACAGAGTGAATGGGGTGTGCAAGATTAAGCTTCCGCGTGC 5136
OY 1741 AAACCTCTCTAAGGAGGAGGATCACTTTCAGGTGTGGGCTCAACCAATATCTGTGTGG 1800
DB 5137 AAACCTCTCTAAGGAGGAGGATCACTTTCAGGTGTGGGCTCAACCAATATCTGTGTGG 5196
```

QY 1801 TCGACGCTCCATGCGAGCCCGAACCAGATGTAGCAGTGTCACTTCCATGTCCACCGAC 1860
 DB 5197 TCGAGCTCCATGCGAGCCCGAACCAGATGTAGCAGTGTCACTTCCATGTCCACCGAC 5256
 QY 1861 CCGTCCCATCATCAGCAGAGACGGCTTAAGCGAGGCTGGCCAGGGGGTCTCCCTCC 1920
 DB 5257 CCGTCCCATCATCAGCAGAGACGGCTTAAGCGAGGCTGGCCAGGGGGTCTCCCTCC 5316
 QY 1921 TTGGCAGCTTTCAGCTAGCCAGTGTCTGCGCCCTTCCGGAAGGCGCATATAC 1980
 DB 5317 TTGGCAGCTTTCAGCTAGCCAGTGTCTGCGCCCTTCCGGAAGGCGCATATAC 5376
 QY 1981 CAAATGACTTCCAGACGCTGCTCATGAGAGCCAACTCTGTGGCGCATGAGATG 2040
 DB 5377 CCGTATGATCTCCCGGACCGCTACCTCATGAGGCCAACTCTGTGGCGCATGAGATG 5436
 QY 2041 GGC 2043
 DB 5437 GGC 5439
 RESULT 15
 ID AAL47276 standard; DNA; 7992 BP.
 AC AAL47276;
 DT 30-AUG-2002 (first entry)
 DE Hepatitis C virus sub-genomic replicon clone 1377-NS3-3'UTR.
 XX Hepatitis; HCV; core-neo; NS3 proteinase/helicase; vaccine; diagnosis;
 KM viricide; hepatotropic; gene therapy; anti-viral; gene; ds.
 XX Hepatitis C virus.
 OS
 XX
 FH Key Location/Qualifiers
 FT CDS /tag= a
 FT /product= "core-neo fusion protein"
 FT CDS 1801..7758
 FT /tag= b
 FT /product= "NS3 proteinase/helicase"
 PN MO200238793-A2.
 XX
 PD 16-MAY-2002.
 XX
 PF 02-NOV-2001; 2001MO-US046350.
 XX
 PR 07-NOV-2000; 2000US-0245866P.
 XX
 PA (ANAD-) ANADYS PHARM INC.
 XX
 PI Bichko V;
 XX
 DR WPI; 2002-490082/52.
 DR P-PSDB; AAO18000; AAO18001.
 PT Novel nucleic acid encoding replication competent recombinant hepatitis C
 PT virus genome useful for screening anti-hepatitis C virus therapeutics and
 PT for vaccine development.
 XX
 PS Claim 6; Page 43-47; 85pp; English.
 CC The present invention provides protein and coding sequences from
 CC Hepatitis C virus (HCV), comprising all or part of the HCV genome and
 CC able to replicate efficiently when transfected into a susceptible cell
 CC line without reducing the growth rate of the cell line by more than 10
 CC fold. The sequences are useful for screening for anti-HCV therapeutics,
 CC for detecting antibodies to HCV in a biological sample such as blood,
 CC serum, plasma, blood cells, lymphocytes, or liver cells from a subject,

CC for deriving authentic HCV components such as replication-complement non-
 CC infectious, replication-defective infection-component, and replication-
 CC defective non-infectious HCV, in gene therapy or gene vaccination
 CC targeted to hepatic tissue for treating an animal infected or susceptible
 CC to HCV infection and for studying HCV infection and propagation. The
 CC present sequence is a clone of a fragment of the HCV genome which encodes
 CC the core-neo and NS3 proteinase/helicase proteins
 XX
 SQ Sequence 7992 BP; 1648 A; 2369 C; 2243 G; 1732 T; 0 U; 0 Other;
 Query Match 87.5%; Score 1787; DB 6; Length 7992;
 Best Local Similarity 92.2%; Pred. No. 0;
 Matches 1883; Conservative 0; Mismatches 160; Indels 0; Gaps 0;
 QY 1 TGGGAGGGGCTTTCACAGGCGCTCACCGGAGATGCCACTTCCTGCCAAACAAG 60
 DB 3397 TGGGAGGGGCTTTCACAGGCGCTCACCGGAGATGCCACTTCCTGCCAAACAAG 3456
 QY 61 CAGGAGAGAGCACTTCCCTTACCTGTGGGCTACAGCTTACTGTGTGGCGCTAGGGCC 120
 DB 3457 CAGGAGAGAGCACTTCCCTTACCTGTGGGCTACAGCTTACTGTGTGGCGCTAGGGCC 3516
 QY 121 CAGGCGCCACTTCATCATGAGTAATGTGAAAGTGTCTCATACGCTAAAGCTTACT 180
 DB 3517 CAGGCTCCACTTCATCATGAGTAATGTGAAAGTGTCTCATACGCTAAAGCTTACT 3576
 QY 181 CTGGCGGGGCGCAACCCCTTGTGTATAGGCTGGGAGCCGCTCCAAAGAGGTCACCCCTC 240
 DB 3577 CTGGCGGGGCGCAACCCCTTGTGTATAGGCTGGGAGCCGCTCCAAAGAGGTCATACCC 3636
 QY 241 ACACACCCCATTAACAAATTCATCATGAGTATGACGAGCTGAGAGGTCGTACAG 300
 DB 3637 ACACACCCCATTAACAAATTCATCATGAGTATGACGAGCTGAGAGGTCGTACAG 3696
 QY 301 AGCACCCTGGTCTGGTGGCGGGGCTCTTCAAGCTCTGGCTGTATGCTTGACACA 360
 DB 3697 AGCACCCTGGTCTGGTGGCGGGGCTCTTCAAGCTCTGGCTGTATGCTTGACACA 3756
 QY 361 GGCAGCGTGTATGTGTGTAGATCATCTGTCCGGCGGCGCGCTATTTTCCCGAC 420
 DB 3757 GGCAGCGTGTATGTGTGTAGATCATCTGTCCGGCGGCGCGCTATTTTCCCGAC 3816
 QY 421 AGGGAAGTCTCTTACAGAGATTCATGATGATGAAAGTGGCGCTGACCTCCCTTAC 480
 DB 3817 AGGGAAGTCTCTTACAGAGATTCATGATGATGAAAGTGGCGCTGACCTCCCTTAC 3876
 QY 481 ATCGACAGGAGATGACGCTGCGGACAGCTTCAAGCAAAAGCGCTGGTGTGACG 540
 DB 3877 ATCGACAGGAGATGACGCTGCGGACAGCTTCAAGCAAAAGCGCTGGTGTGACG 5936
 QY 541 ACAGCCACCAACCAAGCGAGCGCTGTCCCGTGTGTGAGATCCAGTGGGAGCCCTT 600
 DB 3937 ACAGCCACCAACCAAGCGAGCGCTGTCCCGTGTGTGAGATCCAGTGGGAGCCCTT 3996
 QY 601 GAGACCTTCTGGGCGGAAACATGTGAACTTCATCAGCGGAGTACAGTCTTACAGGCG 660
 DB 3997 GAGACCTTCTGGGCGGAAACATGTGAACTTCATCAGCGGAGTACAGTCTTACAGGCG 4056
 QY 661 TTGTCCACTCTGCTGGGATCCCGGATGATGATGATGATGATGATGATGATGATGAT 720
 DB 4057 TTGTCCACTCTGCTGGGATCCCGGATGATGATGATGATGATGATGATGATGATGAT 4116
 QY 721 ACTAGCCGCTTACCAACCAATCTACCTCTGCTTACATCTCTGGGGGATGGGATGAGC 780
 DB 4117 ACCAGCCGCTTACCAACCAATCTACCTCTGCTTACATCTCTGGGGGATGGGATGAGC 4176
 QY 781 GCCCACTGCTCCCGGAGTGTGCTTCACTTGTGAGGCGCGGCTTGTGATGATGATGATGAT 840
 DB 4177 GCCCACTGCTCCCGGAGTGTGCTTCACTTGTGAGGCGCGGCTTGTGATGATGATGATGAT 4236
 QY 841 GCTGTGGAGCATAGGCTTGGGAGGCTTGTGATGATGATGATGATGATGATGATGATGATGAT 900
 DB 4237 GCTGTGGAGCATAGGCTTGGGAGGCTTGTGATGATGATGATGATGATGATGATGATGATGAT 4296

QY 901 GGAGTGGCAGGCGCGCTCGTGGCCCTTTAAGTTCATGAGCGCGCAATGCGCTTCCACCGAG 960
| | | | |
Db 4297 GGGGTGGCAGGCGCGCTCGTGGCCCTTTAAGTTCATGAGCGCGCAATGCGCTTCCACCGAG 4356
QY 961 GACCTGGTTAACTTACTCCCTGCGCATCTCTCTCCCTGGTCCCTGGTCCCTGGGCTGCG 1020
| | | | |
Db 4357 GACCTGGTTAACTTACTCCCTGCGCATCTCTCTCCCTGGTCCCTGGTCCCTGGGCTGCG 4416
QY 1021 TCGCAGCGCATCTGCGTGGGACGCGGCTCCAGGGGAGGGGAGCTGTGCGATGATGAAC 1080
| | | | |
Db 4417 TCGCAGCGCATCTGCGTGGGACGCGGCTCCAGGGGAGGGGAGCTGTGCGATGATGAAC 4476
QY 1081 CGGCTGATAGCGTTCGCTCGCGGGGTAAACAATGTTTCCCGACGCACTATGTGCCAGAG 1140
| | | | |
Db 4477 CGGCTGATAGCGTTCGCTCGCGGGGTAAACAAGTCTCCCGACGCACTATGTGCCAGAG 4536
QY 1141 AGCGAGCGCGCAGCAGCTGTCTCATGATCTCTCCGACCTTACTATCAACCACTGTTG 1200
| | | | |
Db 4537 AGCGAGCGCGCAGCAGCTGTCTCATGATCTCTCTCAAGTCTTCACTCAGCTGCTG 4596
QY 1201 AAGAGGCTCCACGAGGTGAATTAACGAGGACTGCTCCAGCGCCCTGCTCCGGCTCGTGGCTA 1260
| | | | |
Db 4597 AAGAGGCTCCACGAGGTGAATTAACGAGGACTGCTCCAGCGCCCTGCTCCGGCTCGTGGCTA 4656
QY 1261 AGGAGTGTGGGACTGATATGCAAGATTTTGGCTGACTTCAAGAAGCTGGCTCCAGTCC 1320
| | | | |
Db 4657 AGAGATGTTGGGATTTGATATGCAAGTGTGACTGATTTTCAAGAGCTGGCTCCAGTCC 4716
QY 1321 AAGCTTCCTGCGCGGATTAACGGGAGTCCCTTTTCTCATGCCAAGTGGGTCAAGAGGG 1380
| | | | |
Db 4717 AAGCTTCCTGCGCGGATTTGCGGAGTCCCTTTTCTCATGTCMAAGTGGGTCAAGAGGG 4776
QY 1381 GTCTGCGGGGAGAGCGGATTCATGACAGCACCTGCTCATGTGAGACAGATCACCGGA 1440
| | | | |
Db 4777 GTCTGCGGGGAGAGCGGATTCATGACAGCACCTGCTCATGTGAGACAGATCACCGGA 4836
QY 1441 CATGTCAAAAACGTTTCATGAGGATCGTTGGGCTTAAGACCTGTATGAACATGTGCGAT 1500
| | | | |
Db 4837 CATGTCAAAAACGTTTCATGAGGATCGTTGGGCTTAAGACCTGTATGAACATGTGCGAT 4896
QY 1501 GGAACATTTCCCATCAAGCATACACCAAGGCGCCCTGCAAGCGCTCCCGAGCGCAAC 1560
| | | | |
Db 4897 GGAACATTTCCCATCAAGCATACACCAAGGCGCCCTGCAAGCGCTCCCGAGCGCAAC 4956
QY 1561 TATTCAGGGCGCTGTGGCGGGTGTGCTGAGAGATGATGAGAGTTACGCGGGTGGG 1620
| | | | |
Db 4957 TATTCAGGGCGCTGTGGCGGGTGTGCTGAGAGATGATGAGAGTTACGCGGGTGGG 5016
QY 1621 GATTTCCACTACGTGACGAGCATGACCACTGACCAAGTAAATGCGCGTGCAGGTTCCA 1680
| | | | |
Db 5017 GATTTCCACTACGTGACGAGCATGACCACTGACCAAGTAAATGCGCGTGCAGGTTCCA 5076
QY 1681 GCGCCCGGAATTTCTTCAAGAAAGTGGGCTGCGGCTGCAAGGTAAGTCCGCGCTGC 1740
| | | | |
Db 5077 GCGCCCGGAATTTCTTCAAGAAAGTGGGCTGCGGCTGCAAGGTAAGTCCGCGCTGC 5136
QY 1741 AAACCTCTCTACGAGGAGGATCAATTCAGGTGCGGGCTCAACCAATACCTGGTTGGG 1800
| | | | |
Db 5137 AAACCTCTCTACGAGGAGGATCAATTCAGGTGCGGGCTCAACCAATACCTGGTTGGG 5196
QY 1801 TCGCAGCTCCCATGCGAGCCCGAACCAGATGTAGCAGTCTCATTTCCATGTCCACCGAG 1860
| | | | |
Db 5197 TCGCAGCTCCCATGCGAGCCCGAACCAGATGTAGCAGTCTCATTTCCATGTCCACCGAG 5256
QY 1861 CCTTCCCATATCAACAGAGAGAGGCTAAGCGAGGCTGGCCAGGGGGTCTCCCGCTCC 1920
| | | | |
Db 5257 CCTTCCCATATCAACAGAGAGAGGCTAAGCGAGGCTGGCCAGGGGAATCTCCCGCTCC 5316
QY 1921 TTGGCCAGGCTCTTCAAGTACGAGTGTGCGCTTCTCCGAAAGGAGCATATATTACC 1980
| | | | |
Db 5317 TTGGCCAGGCTCTTCAAGTACGAGTGTGCGCTTCTCCGAAAGGAGCATATATTACC 5376

QY 1981 CAAATGACTTCCAGAGCGCTGACCTCATCGAGGCCAACCCTTGTGGCGGAGAGATG 2040
| | | | |
Db 5377 CGTCAATGACTCCCGAGAGCTGACCTCATCGAGGCCAACCCTTGTGGCGGAGAGATG 5436
QY 2041 GGC 2043
| | |
Db 5437 GGC 5439

Search completed: February 25, 2005, 03:32:08
Job time : 964 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2005, 03:10:04 ; Search time 366 Seconds
(without alignments)
9133.643 Million cell updates/sec

Title: US-09-664-363-20

Perfect score: 2043

Sequence: 1 TGGGCGGCGCTTCACAGC.....TGTGCGCGCATGATGCGC 2043

Scoring table:

Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA: *
1: /cgn2_6/prodata/1/ina/5A_COMB.seq: *
2: /cgn2_6/prodata/1/ina/5B_COMB.seq: *
3: /cgn2_6/prodata/1/ina/6A_COMB.seq: *
4: /cgn2_6/prodata/1/ina/6B_COMB.seq: *
5: /cgn2_6/prodata/1/ina/PCITUS_COMB.seq: *
6: /cgn2_6/prodata/1/ina/backfile1.seq: *

Pred: No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|------------------|
| 1 | 2043 | 100.0 | 2043 | 3 | US-08-191-160-20 |
| 2 | 2043 | 100.0 | 3750 | 1 | US-08-191-160-22 |
| 3 | 1796.6 | 87.9 | 7863 | 1 | US-08-324-977-35 |
| 4 | 1796.6 | 87.9 | 7863 | 2 | US-08-324-977-35 |
| 5 | 1796.6 | 87.9 | 7863 | 2 | US-08-324-977-35 |
| 6 | 1796.6 | 87.9 | 7863 | 2 | US-08-324-977-35 |
| 7 | 1796.6 | 87.9 | 7863 | 2 | US-08-324-977-35 |
| 8 | 1796.6 | 87.9 | 7863 | 2 | US-08-324-977-35 |
| 9 | 1796.6 | 87.9 | 7863 | 2 | US-08-324-977-35 |
| 10 | 1796.6 | 87.9 | 7863 | 2 | US-08-324-977-35 |
| 11 | 1796.6 | 87.9 | 7863 | 2 | US-08-324-977-35 |
| 12 | 1796.6 | 87.9 | 7863 | 2 | US-08-324-977-35 |
| 13 | 1796.6 | 87.9 | 7863 | 2 | US-08-324-977-35 |
| 14 | 1796.6 | 87.9 | 7863 | 2 | US-08-324-977-35 |
| 15 | 1796.6 | 87.9 | 7863 | 2 | US-08-324-977-35 |
| 16 | 1796.6 | 87.9 | 7863 | 2 | US-08-324-977-35 |
| 17 | 1796.6 | 87.9 | 7863 | 2 | US-08-324-977-35 |
| 18 | 1796.6 | 87.9 | 7863 | 2 | US-08-324-977-35 |
| 19 | 1796.6 | 87.9 | 7863 | 2 | US-08-324-977-35 |
| 20 | 1796.6 | 87.9 | 7863 | 2 | US-08-324-977-35 |
| 21 | 1796.6 | 87.9 | 7863 | 2 | US-08-324-977-35 |
| 22 | 1796.6 | 87.9 | 7863 | 2 | US-08-324-977-35 |
| 23 | 1796.6 | 87.9 | 7863 | 2 | US-08-324-977-35 |
| 24 | 1796.6 | 87.9 | 7863 | 2 | US-08-324-977-35 |
| 25 | 1796.6 | 87.9 | 7863 | 2 | US-08-324-977-35 |
| 26 | 1796.6 | 87.9 | 7863 | 2 | US-08-324-977-35 |
| 27 | 1796.6 | 87.9 | 7863 | 2 | US-08-324-977-35 |

| | | | | | | |
|----|--------|------|-------|---|-------------------|--------------------|
| 28 | 1783.8 | 87.3 | 8638 | 4 | US-10-029-907-6 | Sequence 6, Appl1 |
| 29 | 1783.8 | 87.3 | 8643 | 4 | US-10-029-907-4 | Sequence 4, Appl1 |
| 30 | 1782.2 | 87.2 | 8638 | 4 | US-10-029-907-7 | Sequence 7, Appl1 |
| 31 | 1782.2 | 87.2 | 8638 | 4 | US-10-029-907-25 | Sequence 25, Appl1 |
| 32 | 1782.2 | 87.2 | 8648 | 4 | US-10-029-907-5 | Sequence 5, Appl1 |
| 33 | 1780.6 | 87.2 | 8001 | 4 | US-09-539-601-16 | Sequence 16, Appl1 |
| 34 | 1780.6 | 87.2 | 8001 | 4 | US-09-539-601-22 | Sequence 22, Appl1 |
| 35 | 1780.6 | 87.2 | 11076 | 4 | US-09-539-601-19 | Sequence 19, Appl1 |
| 36 | 1780.6 | 87.2 | 11076 | 4 | US-09-539-601-25 | Sequence 25, Appl1 |
| 37 | 1777.4 | 87.0 | 8001 | 4 | US-09-539-601-28 | Sequence 28, Appl1 |
| 38 | 1777.4 | 87.0 | 11076 | 4 | US-09-539-601-31 | Sequence 31, Appl1 |
| 39 | 1756.6 | 86.0 | 9595 | 3 | US-09-014-416-4 | Sequence 4, Appl1 |
| 40 | 1748.6 | 85.6 | 9413 | 4 | US-09-827-688-6 | Sequence 6, Appl1 |
| 41 | 1719.8 | 84.2 | 9472 | 4 | US-08-150-2048-96 | Sequence 96, Appl1 |
| 42 | 1345.4 | 65.9 | 9401 | 5 | PCR-US91-02225-9 | Sequence 9, Appl1 |
| 43 | 1343.8 | 65.8 | 6785 | 3 | US-08-444-818-65 | Sequence 65, Appl1 |
| 44 | 1343.8 | 65.8 | 8316 | 3 | US-08-444-818-88 | Sequence 88, Appl1 |
| 45 | 1343.8 | 65.8 | 8987 | 3 | US-08-444-818-137 | Sequence 137, App |

ALIGNMENTS

RESULT 1
US-08-191-160-20
; Sequence 20, Application US/08191160
; Patent No. 6210675
; GENERAL INFORMATION:
; APPLICANT: Highfield, Peter Edmund
; APPLICANT: Rodgers, Brian Collin
; APPLICANT: Tedder, Richard Seton
; APPLICANT: Barbara, John Anthony James
; TITLE OF INVENTION: Viral Agent
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Rothwell, Pigs, Ernst & Kurz
; STREET: 1700 K Street
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy diskette, 5.25 inch, 360 Kb storage
; COMPUTER: IBM AT compatible
; OPERATING SYSTEM: MS-DOS V3.2
; SOFTWARE: Wordperfect 5.0 (DOS text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08191,160
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/628,516
; FILING DATE: 17 DEC 1990
; APPLICATION NUMBER: UK 89 28 562.1
; FILING DATE: 18 DEC 1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: UK 90 04 414.0
; FILING DATE: 27 FEB 1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: UK 90 04 814.1
; FILING DATE: 03 MAR 1990
; ATTORNEY/AGENT INFORMATION:
; NAME: F. Anthony Pigg
; REGISTRATION NUMBER: 27,195
; REFERENCE/DOCKET NUMBER: 1645-103A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 833-5740
; TELEFAX: (202) 833-5744
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2043 base pairs
; TYPE: nucleotide with corresponding protein
; STRANDEDNESS: single

```

?      TOPOLOGY: linear
?      MOLECULE TYPE: CDNA to genomic RNA
?      ORIGINAL SOURCE:
?      ORGANISM: human; serum infectious for PT-NANBH
?      IMMEDIATE SOURCE:
?      LIBRARY: clone 156/92
?      FEATURE:
?      LOCATION: from 1 to 2043 bp portion of the PT-NANBH
?      LOCATION: polyprotein
?      OTHER INFORMATION: probably encodes viral non-structural
?      OTHER INFORMATION: proteins
?      OS-08-191-160-20

```

| | | | | |
|----------------------------|---------|--------------|-------|--------------|
| Query Match | 100.0% | Score 2043; | DB 3; | Length 2043; |
| Best Local Similarity | 100.0%; | Pred. No. 0; | | |
| Matches 2043; Conservative | 0; | Mismatches | 0; | Gaps 0; |

[illegible]

| | | | |
|----|------|---|------|
| QY | 841 | UCTGTGGCAGCATAGGCGCTTTGGGAAGGTCCTTTGGACATCTTTGGGGGCTATAGAGCA | 900 |
| Db | 841 | GCTGTGGCAGCATAGGCGCTTTGGGAAGGTCCTTTGGACATCTTTGGGGGCTATAGAGCA | 900 |
| QY | 901 | GGAGTGGCAGGCGGCGCTCGTGGCCTTTAAAGTCATGAGCGGCAAAATGCCCTCCACCGAG | 960 |
| Db | 901 | GGAGTGGCAGGCGGCGCTCGTGGCCTTTAAAGTCATGAGCGGCAAAATGCCCTCCACCGAG | 960 |
| QY | 961 | GACCTGGTTAACTTACTCCCTGGCCATCTCTCTCTGGTGGCCCTGGTCTGTGGGGGTCTGT | 1020 |
| Db | 961 | GACCTGGTTAACTTACTCCCTGGCCATCTCTCTCTGGTGGCCCTGGTCTGTGGGGGTCTGT | 1020 |
| QY | 1021 | TGCCGACGCACTACTGCGCTCGGCAAGTGGGTCCAGAGGGAGGGGGCTGTGCAGTGGATGAAC | 1080 |
| Db | 1021 | TGCCGACGCACTACTGCGCTCGGCAAGTGGGTCCAGAGGGAGGGGGCTGTGCAGTGGATGAAC | 1080 |
| QY | 1081 | CGGCTGATAGCGTTGCGCTCGCGGGGGTAAACATGTTTCCCGCAGCACTATGTGGCAGAG | 1140 |
| Db | 1081 | CGGCTGATAGCGTTGCGCTCGCGGGGGTAAACATGTTTCCCGCAGCACTATGTGCCAGAG | 1140 |
| QY | 1141 | AGCGACGCGCGCAGCAGCTGTCTACTCAGATCCYCTCCGACCTTACTATCACCCAACTGTGG | 1200 |
| Db | 1141 | AGCGACGCGCGCAGCAGCTGTCTACTCAGATCCYCTCCGACCTTACTATCACCCAACTGTGG | 1200 |
| QY | 1201 | AAGAGGCTCCACCACTGGATTTAAAGAGAGACTGCTCCAGCCCTGCTCCGGCTGTGGCTA | 1260 |
| Db | 1201 | AAGAGGCTCCACCACTGGATTTAAAGAGAGACTGCTCCAGCCCTGCTCCGGCTGTGGCTA | 1260 |
| QY | 1261 | AGGAGTGTGGGGAGCTGGATATGACAGTTTGGCTGACTTCAAGACCTGGAGTCCAGTCC | 1320 |
| Db | 1261 | AGGAGTGTGGGGAGCTGGATATGACAGTTTGGCTGACTTCAAGACTTGGCTCAGTCC | 1320 |
| QY | 1321 | AAGCTCTGCGCGCAGTTACCGGGAGTCCCTTTTCTCATGCCAACTGTGGGTACAAGGGG | 1380 |
| Db | 1321 | AAGCTCTGCGCGCAGTTACCGGGAGTCCCTTTTCTCATGCCAACTGTGGGTACAAGGGG | 1380 |
| QY | 1381 | GTCTGGCGGGAGACGGGATCATGCAACAACCTGCTCATGTGGAGACAGATCACCGGA | 1440 |
| Db | 1381 | GTCTGGCGGGAGACGGGATCATGCAACAACCTGCTCATGTGGAGACAGATCACCGGA | 1440 |
| QY | 1441 | CATGCAAAAAAGTTCCATGAGGATGGTGGGGCTGAAGACCTGTACTTAACATGTGGCAT | 1500 |
| Db | 1441 | CATGCAAAAAAGGTTCCATGAGGATGTTGGGCTGAAGACCTGTACTTAACATGTGGCAT | 1500 |
| QY | 1501 | GGAACATTTCCCATCAACGCAATACCAACCGGGCCCTGCAAGCCCTCCCAAGCGCCAAC | 1560 |
| Db | 1501 | GGAACATTTCCCATCAACGCAATACCAACCGGGCCCTGCAAGCCCTCCCAAGCGCCAAC | 1560 |
| QY | 1561 | TATTCGAGGGCGCTGTGGCGGGGTGGCTGTGAGGAGTACGTGGAGGTTTAAAGCGGGTGGGG | 1620 |
| Db | 1561 | TATTCGAGGGCGCTGTGGCGGGGTGGCTGTGAGGAGTACGTGGAGGTTTAAAGCGGGTGGGG | 1620 |
| QY | 1621 | GATTTCCACTACGTGAGAGGATGACCACTGACAAAGTAAATGCCCCGTGCAGGTTCCA | 1680 |
| Db | 1621 | GATTTCCACTACGTGAGAGGATGACCACTGACAAAGTAAATGCCCCGTGCAGGTTCCA | 1680 |
| QY | 1681 | GGCCCGCAATCTTCAACAAGATGGATGGGGTGGCTGTGACAGGTACGCTCGCGCGCTGC | 1740 |
| Db | 1681 | GGCCCGCAATCTTCAACAAGATGGATGGGGTGGCTGTGACAGGTACGCTCGCGCGCTGC | 1740 |
| QY | 1741 | AAACCTCTCTACCGGAGAGAGTCAATTCAGGTGGGGCTCAACCAATACCTGTGTGGG | 1800 |
| Db | 1741 | AAACCTCTCTACCGGAGAGAGTCAATTCAGGTGGGGCTCAACCAATACCTGTGTGGG | 1800 |
| QY | 1801 | TGCGACGCTCCATGCGAGCCCGAAACCGGATGTAGCAATGCTCACTTCCATGTCCACCGAC | 1860 |
| Db | 1801 | TGCGACGCTCCATGCGAGCCCGAAACCGGATGTAGCAATGCTCACTTCCATGTCCACCGAC | 1860 |
| QY | 1861 | CCCTGCCACATCAAGCAGAGAGCGGCTAAGCGCAGGCTGGGCGAGGGGGTCTCCCGCTCC | 1920 |
| Db | 1861 | CCCTGCCACATCAAGCAGAGAGCGGCTAAGCGCAGGCTGGGCGAGGGGGTCTCCCGCTCC | 1920 |
| QY | 1921 | TTGGCAGGCTTTACGCTAGCAGGTCAGTTGTCTGGCCCTTCTCGAAGCGGACATACATTACC | 1980 |

|||||
Db 1921 TTGGCCAGCTCTTCAAGTACGACGATGTCGCGCTTCTCCGACAGGCGACATACATTACC 1980
QY 1981 CAAATGACTCTCCGAGAGCTGACCTCAACGAGGCGCAACCTCTGGGCGGATGAGATG 2040
Db 1981 CAAATGACTCTCCGAGAGCTGACCTCAACGAGGCGCAACCTCTGGGCGGATGAGATG 2040
QY 2041 GGC 2043
Db 2041 GGC 2043
RESULT 2
US-08-191-160-22
Sequence 22, Application US/08191160
Patent No. 6210675
GENERAL INFORMATION:
APPLICANT: Highfield, Peter Edmund
APPLICANT: Rodgers, Brian Colin
APPLICANT: Redder, Richard Seton
APPLICANT: Barbara, John Anthony James
TITLE OF INVENTION: Viral Agent
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSER: Rothwell, F199, Ernst & Kurz
STREET: 1700 K Street
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy diskette, 5.25 inch, 360 Kb storage
COMPUTER: IBM AT compatible
OPERATING SYSTEM: MS-DOS V3.2
SOFTWARE: Wordperfect 5.0 (DOS text)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/191,160
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/628,516
FILING DATE: 17 DEC 1990
APPLICATION NUMBER: UK 89 28 562.1
FILING DATE: 18 DEC 1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: UK 90 04 414.0
FILING DATE: 27 FEB 1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: UK 90 04 814.1
FILING DATE: 03 MAR 1990
ATTORNEY/AGENT INFORMATION:
NAME: E. Anthony F199
REGISTRATION NUMBER: 27,195
REFERENCE/DOCKET NUMBER: 1645-103A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 833-5740
TELEFAX: (202) 833-5744
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 3750 base pairs
TYPE: nucleotide with corresponding protein
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to genomic RNA
ORIGINAL SOURCE:
ORGANISM: human; serum infectious for PT-NANBH
IMMEDIATE SOURCE:
LIBRARY: cDNA clones from 3' end of the genome
FEATURE:
LOCATION: from 1 to 3750 bp portion of the PT-NANBH
OTHER INFORMATION: viral non-structural proteins
US-08-191-160-22

Query Match 100.0%; Score 2043; DB 3; Length 3750;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2043; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGGAGGGGCGTCTTCAACAGGCTTCAACCAAGTGAATGCCACTTCTGTCCAAACAAAG 60
Db 1 TGGAGGGGCGTCTTCAACAGGCTTCAACCAAGTGAATGCCACTTCTGTCCAAACAAAG 60
QY 61 CAGGAGAGAGCAACTTCCCTACCTGAGGCTGACAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 120
Db 61 CAGGAGAGAGCAACTTCCCTACCTGAGGCTGACAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 120
QY 121 CAGGCCCCACCTTCATCATGAGATCAAAATGGAAGTCTCATACGGCTAAAGCTTACT 180
Db 121 CAGGCCCCACCTTCATCATGAGATCAAAATGGAAGTCTCATACGGCTAAAGCTTACT 180
QY 181 CTGGCGGGGCGCAACACCTTGTGTATAGGCTGGAGCCGTCCAAACAGAGTCAACCTTC 240
Db 181 CTGGCGGGGCGCAACACCTTGTGTATAGGCTGGAGCCGTCCAAACAGAGTCAACCTTC 240
QY 241 ACACACCCCATPACCAATTCATCATGAGATGCAATGTCAGCCGACCTGAGGTGCTACG 300
Db 241 ACACACCCCATPACCAATTCATCATGAGATGCAATGTCAGCCGACCTGAGGTGCTACG 300
QY 301 AGCACCTGGGTGCTGTGGGCGGGGCTCTTGACGCTGCGTGTATTTGCTTACCAACA 360
Db 301 AGCACCTGGGTGCTGTGGGCGGGGCTCTTGACGCTGCGTGTATTTGCTTACCAACA 360
QY 361 GGCAGCGTGTATCTTGTGGTATGATCATCTTGTCCGGGCGGCGCTATTGTTCCGAC 420
Db 361 GGCAGCGTGTATCTTGTGGTATGATCATCTTGTCCGGGCGGCGCTATTGTTCCGAC 420
QY 421 AGGGAAGTCTCTTACCAAGAGTTCATGATGGAAGTGGCGCTCCACCTCCCTTAC 480
Db 421 AGGGAAGTCTCTTACCAAGAGTTCATGATGGAAGTGGCGCTCCACCTCCCTTAC 480
QY 481 ATCGAGCAGGGGAATGACGCTGCGGAGGATTCAGCAAAAAGGCTGGGTGTCGAG 540
Db 481 ATCGAGCAGGGGAATGACGCTGCGGAGGATTCAGCAAAAAGGCTGGGTGTCGAG 540
QY 541 ACAGCCACCAAGCAAGCGAGGCGCTGCTCCGCTGTGAGTCCAGTGGCGAGCCCTT 600
Db 541 ACAGCCACCAAGCAAGCGAGGCGCTGCTCCGCTGTGAGTCCAGTGGCGAGCCCTT 600
QY 601 GAGACCTTCTGGGCGAAACACATGTGAATCTTCAACGCGGATACATCTTACAGGC 660
Db 601 GAGACCTTCTGGGCGAAACACATGTGAATCTTCAACGCGGATACATCTTACAGGC 660
QY 661 TTGTCACTGCGCTCGGGAATCCCGGATTCATCATGAGTGGCGGTTCACAGCCTCTGTC 720
Db 661 TTGTCACTGCGCTCGGGAATCCCGGATTCATCATGAGTGGCGGTTCACAGCCTCTGTC 720
QY 721 ACTAGCCGCTCACCAACCAATCTACCTCTGTCTTAAACATCTGGGGGATGAGTACCC 780
Db 721 ACTAGCCGCTCACCAACCAATCTACCTCTGTCTTAAACATCTGGGGGATGAGTACCC 780
QY 781 GCCCACTGCTCCCGAGTGTCTTCAAGTTGTAGGCGCGGCAATGTCTGTGCG 840
Db 781 GCCCACTGCTCCCGAGTGTCTTCAAGTTGTAGGCGCGGCAATGTCTGTGCG 840
QY 841 GCTGTGGCAGATAGGCTTGGGAAGTGTCTTGGACATCTTGGCGGGCTATGAGACA 900
Db 841 GCTGTGGCAGATAGGCTTGGGAAGTGTCTTGGACATCTTGGCGGGCTATGAGACA 900
QY 901 GGAATGAGAGGCGCTGTGGCTTAAAGTATGAGAGCGGGAATGCTTCCACCGAG 960
Db 901 GGAATGAGAGGCGCTGTGGCTTAAAGTATGAGAGCGGGAATGCTTCCACCGAG 960
QY 961 GACCTGTAACTTACTCTCTGCAATCTCTCTCTGAGGCTGTGCTGTGAGGCTGTG 1020
Db 961 GACCTGTAACTTACTCTCTGCAATCTCTCTCTGAGGCTGTGCTGTGAGGCTGTG 1020

| | | | |
|----|------|---|------|
| OY | 1021 | TGCCACGAGTAATGCTGCTGCGACAGTGGGTCCAGGGAGAGGGGGCTGTGCACTGATGAAC | 1080 |
| Db | 1021 | TGCCACGAGTAATGCTGCTGCGACAGTGGGTCCAGGGAGAGGGGGCTGTGCACTGATGAAC | 1080 |
| OY | 1081 | CGGCTGATAGCGTTGCGCTCGCGGGGGTAACCAATGTTTCCCCACGCACTATGTGCGAG | 1140 |
| Db | 1081 | CGGCTGATAGCGTTGCGCTCGCGGGGGTAACCAATGTTTCCCCACGCACTATGTGCGAG | 1140 |
| OY | 1141 | AGCGACGCGCGAGACGCGTGCACCTCAGATCCCTTCGACCTTAACTATCACCCACTGTG | 1200 |
| Db | 1141 | AGCGACGCGCGAGACGCGTGCACCTCAGATCCCTTCGACCTTAACTATCACCCACTGTG | 1200 |
| OY | 1201 | AAGAGGCTCCACCAAGTGAATTAAAGAGAACTGTCCACGCCCTGTCTCGGCTCGTGCCTA | 1260 |
| Db | 1201 | AAGAGGCTCCACCAAGTGAATTAAAGAGAACTGTCCACGCCCTGTCTCGGCTCGTGCCTA | 1260 |
| OY | 1261 | AGGATGTTTTGGGACTGATATGCAACAATTTTGGCTGACTTAAGAACTGTGCTCCAGTC | 1320 |
| Db | 1261 | AGGATGTTTTGGGACTGATATGCAACAATTTTGGCTGACTTAAGAACTGTGCTCCAGTC | 1320 |
| OY | 1321 | AAGCTCTGCGCGCGATTACCGGAGTCCCTTTTCTCATGCAACACTGGGTAACAAGGG | 1380 |
| Db | 1321 | AAGCTCTGCGCGCGATTACCGGAGTCCCTTTTCTCATGCAACACTGGGTAACAAGGG | 1380 |
| OY | 1381 | GTCTGGCGGGAGACGCGCATATGCAACCACTGTCTCATGTGGAGCAAGATCAACGGA | 1440 |
| Db | 1381 | GTCTGGCGGGAGACGCGCATATGCAACCACTGTCTCATGTGGAGCAAGATCAACGGA | 1440 |
| OY | 1441 | CATGTCAAAAACGTTTCATAGAGATCGTTGGGAGCTAAGACCTGTAGTAAATGATGGCAT | 1500 |
| Db | 1441 | CATGTCAAAAACGTTTCATAGAGATCGTTGGGAGCTAAGACCTGTAGTAAATGATGGCAT | 1500 |
| OY | 1501 | GGAACATTCCCATCAACGCAATACCAACGCGGCCCTGACGCGCTCCACGCGCAAAAC | 1560 |
| Db | 1501 | GGAACATTCCCATCAACGCAATACCAACGCGGCCCTGACGCGCTCCACGCGCAAAAC | 1560 |
| OY | 1561 | TATTTCAGAGGCGCTGTGGCGGGTGGCTCTGAGAGTACGTGAGATTAAACCGGGTGGGG | 1620 |
| Db | 1561 | TATTTCAGAGGCGCTGTGGCGGGTGGCTCTGAGAGTACGTGAGATTAAACCGGGTGGGG | 1620 |
| OY | 1621 | GATTTCACACTAGTACGAGCATGACCACTGACAACGTAATAATGCCCTGACAGTTTCA | 1680 |
| Db | 1621 | GATTTCACACTAGTACGAGCATGACCACTGACAACGTAATAATGCCCTGACAGTTTCA | 1680 |
| OY | 1681 | GCCCCCGAATTTCTTCAAGAAAGTGATGAGGGGCGCGCTGCAACAGTTACGCTCCGCGCTG | 1740 |
| Db | 1681 | GCCCCCGAATTTCTTCAAGAAAGTGATGAGGGGCGCGCTGCAACAGTTACGCTCCGCGCTG | 1740 |
| OY | 1741 | AAACCTCTCTACCGGAGGAGAGGTCACTTCCAGGTCGCGGCTCAACCAATACCTGGTTGG | 1800 |
| Db | 1741 | AAACCTCTCTACCGGAGGAGAGGTCACTTCCAGGTCGCGGCTCAACCAATACCTGGTTGG | 1800 |
| OY | 1801 | TCGCAGCTCCCATGCGAGCCCGAACCAGATGTAGAGAGTCTCACTTCATGTCTCACCGAC | 1860 |
| Db | 1801 | TCGCAGCTCCCATGCGAGCCCGAACCAGATGTAGAGAGTCTCACTTCATGTCTCACCGAC | 1860 |
| OY | 1861 | CCCTCCCACTACACGACGAGACCGGCTAAGCCCAAGCTGGCCAGGGGGCTTCCCTCCCTCC | 1920 |
| Db | 1861 | CCCTCCCACTACACGACGAGACCGGCTAAGCCCAAGCTGGCCAGGGGGCTTCCCTCCCTCC | 1920 |
| OY | 1921 | TTGGCGACGCTCTTCAAGCTAGCGAGTTGTCTGTGGCTTCTCTGAAAGCGACATACATTAC | 1980 |
| Db | 1921 | TTGGCGACGCTCTTCAAGCTAGCGAGTTGTCTGTGGCTTCTCTGAAAGCGACATACATTAC | 1980 |
| OY | 1981 | CAAAATGACTTCCACAGACGCTGACCTCAATCGAGGCCAAACCTCTGTGGCGGATAGATG | 2040 |
| Db | 1981 | CAAAATGACTTCCACAGACGCTGACCTCAATCGAGGCCAAACCTCTGTGGCGGATAGATG | 2040 |
| OY | 2041 | GGC 2043 | |
| Db | 2041 | GGC 2043 | |

```

US-08-324-977-35
Sequence 35, Application US/08324977
Patent No. 574739
GENERAL INFORMATION:
APPLICANT: OKAYAMA, Hiroto
APPLICANT: FUKU, Isao
APPLICANT: MORI, Chiato
APPLICANT: TAKAMIZAWA, Akahisa
APPLICANT: YOSHIDA, Iwao
TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC
TYPE OF INVENTION: CDNA AND ANTIGEN POLYPEPTIDE
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: Armstrong, Westerman, Hattori, Mclelland &
ADDRESSER: Naughton
STREET: 1725 K St. N.W. Suite 1000
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 in, 1.44mb
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS, Version 5.0
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/324,977
FILING DATE: 18-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 2-167466
FILING DATE: 25-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 2-230921
FILING DATE: 31-AUG-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 2-305605
FILING DATE: 09-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/099,706
FILING DATE: 30-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/769,996
FILING DATE: 02-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/635,451
FILING DATE: 28-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: Stevens-Smith, Theresa M.
REGISTRATION NUMBER: 36,281
REFERENCE/DOCKET NUMBER: 900703D
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 659-2930
TELEFAX: (202) 887-0357
TELEX: 440142
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 7863 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA from genomic RNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..7863
NAME/KEY: misc feature
LOCATION: 1..7863
OTHER INFORMATION: /note="sequence = 1500 - 9362 of
US-08-324-977-35

```

Best Local Similarity 92.5%; Pred. No. 0; Matches 1869; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

1 TGGAGGGGGCTCTTCAAGAGCCCAACCGATGATGCCACTTCTGTCCTCCCAACAAAG 60
 3505 TGGAGAGTGTCTTCAAGAGCCCAACCGATGATGCCACTTCTGTCCTCCCAACAAAG 3564
 61 CAGGAGAGAGCAACTTCCCTCACTGATGAGCTACAGAGCTACTGTGTGCGCTAGAGGC 120
 3565 CAGGAGAGAGCAACTTCCCTCACTGATGAGCTACAGAGCTACTGTGTGCGCTAGAGGC 3624
 121 CAGGCCCCACTTCATCATGAGATCAAAATGTGAAATGTCTCATACGCTTAAAGCTTACT 180
 3625 CAGGCCCCACTTCATCATGAGATCAAAATGTGAAATGTCTCATACGCTTAAAGCTTACT 3684
 181 CTGGCGGGGGCAACCCCTTGTGTATAGAGCTGGAGCCGTCACAAACAGAGCTACCCCTC 240
 3685 CTGGCGGGGGCAACCCCTTGTGTATAGAGCTGGAGCCGTCACAAATGAGAGCTACCCCTC 3744
 241 ACACACCCCATTAACCAATTCATCATGAGATGATGATGATGATGATGATGATGATGATG 300
 3745 ACCCACCCCATTAACCAATTCATCATGAGATGATGATGATGATGATGATGATGATGATG 3804
 301 AGCAGCTGGATGCTGGTGGGCGGGGCTCTTGAAGCTCTGGCTGCTGATTTGTTGACAA 360
 3805 AGCAGCTGGATGCTGGTGGGCGGGGCTCTTGAAGCTCTGGCTGCTGATTTGTTGACAA 3864
 361 GGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 420
 3865 GGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3924
 421 AGGGAAGTCTCTACAGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 480
 3925 AGGGAAGTCTCTACAGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 3984
 481 ATGAGCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 540
 3985 ATGAGCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4044
 541 ACAGCAGCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 600
 4045 ACAGCAGCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4104
 601 GAGACCTTCTGGGCGGAAACATGATGATGATGATGATGATGATGATGATGATGATGATG 660
 4105 GAGACCTTCTGGGCGGAAACATGATGATGATGATGATGATGATGATGATGATGATGATG 4164
 661 TTGTCACTCTGCTGGAGATCCGCGATTTGATCATGATGATGATGATGATGATGATGATG 720
 4165 TTGTCACTCTGCTGGAGATCCGCGATTTGATCATGATGATGATGATGATGATGATGATG 4224
 721 ACTAGCCGCTCAACCAATTCATCATGATGATGATGATGATGATGATGATGATGATGATG 780
 4225 ACTAGCCGCTCAACCAATTCATCATGATGATGATGATGATGATGATGATGATGATGATG 4284
 781 GCGCACTGCTCTGCTGGAGATCCGCGATTTGATCATGATGATGATGATGATGATGATGATG 840
 4285 GCGCACTGCTCTGCTGGAGATCCGCGATTTGATCATGATGATGATGATGATGATGATGATG 4344
 841 GCGTGTGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 900
 4345 GCGTGTGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4404
 901 GAGATGAG 960
 4405 GAGATGAG 4464
 961 GACCTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1020
 4465 GACCTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4524
 1021 TGGGAGCGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1080

4525 TGTGACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4584
 1081 CGGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1140
 4585 CGGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4644
 1141 AGCAGCGCGAG 1200
 4645 AGCAGCGCGAG 4704
 1201 AAGAGCTTCAACAGAGATTAACAGAGATTAACAGAGATTAACAGAGATTAACAGAGAT 1260
 4705 AAGAGCTTCAACAGAGATTAACAGAGATTAACAGAGATTAACAGAGATTAACAGAGAT 4764
 1261 AAGAGCTTCAACAGAGATTAACAGAGATTAACAGAGATTAACAGAGATTAACAGAGAT 1320
 4765 AAGAGCTTCAACAGAGATTAACAGAGATTAACAGAGATTAACAGAGATTAACAGAGAT 4824
 1321 AAGCTCTGCGCGATTAACAGAGATTAACAGAGATTAACAGAGATTAACAGAGATTAAC 1380
 4825 AAGCTCTGCGCGATTAACAGAGATTAACAGAGATTAACAGAGATTAACAGAGATTAAC 4884
 1381 GTTGGCGGGAG 1440
 4885 GTTGGCGGGAG 4944
 1441 CATGTCAAAAAGGATTCATGAGATGATGAGATGATGAGATGATGAGATGATGAGATG 1500
 4945 CATGTCAAAAAGGATTCATGAGATGATGAGATGATGAGATGATGAGATGATGAGATG 5004
 1501 GGAACATTTCCCATCAACAGATTAACAGAGATTAACAGAGATTAACAGAGATTAACAGAG 1560
 5005 GGAACATTTCCCATCAACAGATTAACAGAGATTAACAGAGATTAACAGAGATTAACAGAG 5064
 1561 TATTCAGGCGCTGTGGGCTGTGGGCTGTGGGCTGTGGGCTGTGGGCTGTGGGCTGTGGG 1620
 5065 TATTCAGGCGCTGTGGGCTGTGGGCTGTGGGCTGTGGGCTGTGGGCTGTGGGCTGTGGG 5124
 1621 GATTTCCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1680
 5125 GATTTCCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5184
 1681 GCCCCGGAATTTCTTCAAGAGATGATGATGATGATGATGATGATGATGATGATGATG 1740
 5185 GCTCTGAATTTCTTCTGAGAGATGATGATGATGATGATGATGATGATGATGATGATG 5244
 1741 AAACCTTCTTCAAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATG 1800
 5245 AAGGCTTCTTCAAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATG 5304
 1801 TGGAGCTTCCATGAG 1860
 5305 TCAAGAGCTTCAATGAG 5364
 1861 CCTCCCATTAACAG 1920
 5365 CCTCCCATTAACAG 5424
 5425 TTGGCAGCTTCTTCAAG 5484
 1981 CAAATGATTTCCCATGAG 2040
 5485 CACATGATTTCTTCCGAG 5544
 2041 GGC 2043
 5545 GGC 5547

RESULT 4
 US-08-384-616-35

```

1Sequence 35, Application US/08384616
2Patent No. 5847101
3GENERAL INFORMATION:
4APPLICANT: OKAYAMA, Hiroto
5APPLICANT: FUKU, Isao
6APPLICANT: MORI, Chisato
7APPLICANT: TAKAMIZAWA, Akahisa
8APPLICANT: YOSHIDA, Iwao
9TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC
10TITLE OF INVENTION: cDNA AND ANTIGEN POLYPEPTIDE
11NUMBER OF SEQUENCES: 50
12CORRESPONDENCE ADDRESS:
13ADDRESS: Armstrong, Westernman, Hatlori, McIeland &
14ADDRESS: Naughton
15STREET: 1725 K St. N.W. Suite 1000
16CITY: Washington
17STATE: D.C.
18COUNTRY: U.S.A.
19ZIP: 20006
20COMPUTER READABLE FORM:
21MEDIUM TYPE: Diskette, 3.5 in, 1.44MB
22COMPUTER: IBM PC compatible
23OPERATING SYSTEM: PC-DOS/MS-DOS, Version 5.0
24SOFTWARE: ASCII
25CURRENT APPLICATION DATA:
26APPLICATION NUMBER: US/08/384,616
27FILING DATE:
28CLASSIFICATION: 424
29PRIOR APPLICATION DATA:
30APPLICATION NUMBER: US 07/769,996
31FILING DATE: 02-OCT-1991
32APPLICATION NUMBER: JP 2-167466
33FILING DATE: 25-JUN-1990
34PRIOR APPLICATION DATA:
35APPLICATION NUMBER: JP 2-230921
36FILING DATE: 31-AUG-1990
37PRIOR APPLICATION DATA:
38APPLICATION NUMBER: JP 2-305605
39FILING DATE: 09-NOV-1990
40PRIOR APPLICATION DATA:
41APPLICATION NUMBER: US 07/635,451
42FILING DATE: 28-DEC-1990
43ATTORNEY/AGENT INFORMATION:
44NAME: Stevens-Smith, Theresa M.
45REGISTRATION NUMBER: 36,281
46REFERENCE/DOCKET NUMBER: 900703B
47TELECOMMUNICATION INFORMATION:
48TELEPHONE: (202) 659-2293
49TELEFAX: (202) 887-0357
50INFORMATION FOR SEQ ID NO: 35:
51SEQUENCE CHARACTERISTICS:
52LENGTH: 7863 base pairs
53TYPE: nucleic acid
54STRANDEDNESS: single
55TOPOLOGY: linear
56MOLECULE TYPE: cDNA from genomic RNA
57FEATURE:
58NAME/KEY: CDS
59LOCATION: 1..7863
60FEATURE:
61NAME/KEY: misc feature
62LOCATION: 1..7863
63OTHER INFORMATION: /note="sequence = 1500 - 9362 of
64US-08-384-616-35

```

| | Query Match | 87.9%; | Score 1796.6; | DB 2; | Length 7863; |
|----|----------------------------|--|-----------------|-----------|--------------|
| | Best local Similarity | 92.5%; | Pred. No. 0; | | |
| | Matches 1889; Conservative | | Mismatches 154; | Indels 0; | Gaps 0; |
| OY | 1 | TTGGAGGGCGTCTTTCACAGGCTTACACCCACGTGATGCGCACTTCTGTGCCAAACAAAG | 60 | | |

| | | | |
|----|------|---|------|
| Db | 3505 | TTGGAGAGTGTCTTACACAGAGCTCACCCCATATAGATGACACACTCTTGTGCCAGACCAAG | 3564 |
| OY | 61 | CAGGACGAGACAACTTCCCTTACCTGCTGGCGTACCAAGGTACTGTGTGGCTAAGGGCC | 120 |
| Db | 3565 | CAGGCAGAGACAACTTCCCTTACCTGCTGGTACATACCAAGCAAGGTGTGGCCAGGGCT | 3624 |
| OY | 121 | CAGGCCCACTCCATCATATGGAGTCAATATGGAAAGTGTCTATACGGGCTTAAAGCTACT | 180 |
| Db | 3625 | CAGGCCCACTCCATCATATGGAGTCAATATGGAAAGTGTCTATACGGGCTTAAAGCTACT | 3684 |
| OY | 181 | CTGGCGGGGCCAACCCCTTGGTGTATAGCTGGAGACCGTCCAAAACGAGGTCAACCTC | 240 |
| Db | 3685 | CTGCACGGGGCCAAACCCCTTGGTGTATACAGGCTGGAGACCGTCCGAATAGAGTCAACCTC | 3744 |
| OY | 241 | ACACACCCCATTAACCAATTCATATGAGCATGATCTACGCCACCTGAGAGTGTCAAG | 300 |
| Db | 3745 | ACCCACCCCATTAACCAATTCATATGAGCATGATCTACGCCACCTGAGAGTGTCAAG | 3804 |
| OY | 301 | AGCACCTGGGGGCTGGTGGGGGGGGGCTTGAAGCTCGCTGGGTATGCTTACAAACA | 360 |
| Db | 3805 | AGCACCTGGGGGCTGGTGGGGGGGGGCTTGGACACTCGCTGGGTATGCTTACAAACA | 3864 |
| OY | 361 | GGCAGCGGTGCATTGTGGGTAGAGATCATCTTGTCCGGGCGCGGCTATATGTTCCGAC | 420 |
| Db | 3865 | GGCAGGTGTGCATTGTGGGTAGAGATCATCTTGTCCGGGAGCGGCACTTGTTCGAC | 3924 |
| OY | 421 | AGGGAAGTCTCTAACAGAGATTGATGAGATGGAAGATGCGCGCTGCACCTTCCCTTAC | 480 |
| Db | 3925 | AGGGAAGTCTCTAACAGAGATTGATGAGATGGAAGATGCGCGCTGCACCTTCCCTTAC | 3984 |
| OY | 481 | ATCGACGACGGAATATGCAAGCTTCGCGGACAGTTTCAAGAAAAGGCTCGGGTTGCTGAG | 540 |
| Db | 3985 | ATCGACGACGGAATATGCAAGCTTCGCGGACAGTTTCAAGAAAAGGCTCGGGTTGCTGAA | 4044 |
| OY | 541 | ACAGCCACCAAGCAAGCGAGGCGGCTGCTCCGCTGGTGGAGTCCAAGTGGCGAGCCCTT | 600 |
| Db | 4045 | ACAGCCACCAAGCAAGCGAGGCGGCTGCTCCGCTGGTGGAGTCCAAGTGGCGAGCCCTT | 4104 |
| OY | 601 | GAGACCTTCTGGGGGAAACAACATGTGAACTTTCATACGCGGAGTACAGTACTTATAGACGC | 660 |
| Db | 4105 | GAGACCTTCTGGGGGAAACAACATGTGAACTTTCATACGCGGAGTACAGTACTTATAGACGC | 4164 |
| OY | 661 | TTGTTCACACTCTGCTGGGGAATCCCGCATTTGCATCACTGATGGGTTTACAGCCTTGTCT | 720 |
| Db | 4165 | TTATTCACACTCTGCTGGGGAATCCCGCAATACATCATTTGATGGGATTCACAGCCTTATTC | 4224 |
| OY | 721 | ACTAGCCCGCTCACACCCCAATTCACCTCGCTTAACATCTGGGGGGAATGGGATACC | 780 |
| Db | 4225 | ACAGACCCCGCTCACACCCCAATTCACCTCGCTTAACATCTGGGGGGAATGGGATACC | 4284 |
| OY | 781 | GCCCAACTCGCTCCCGCCCAAGTGTCTTCAGACTTTCGTATGGCGCGCGCACTTCTGTGTGCG | 840 |
| Db | 4285 | GCCCAACTCGCGCCCGCCCAAGTGTCTTCGTATGGCGCGCGCGCATGCGCGGTGCG | 4344 |
| OY | 841 | GCTGTGGACGATAGAGCTTTGGGAAGGTGCTTGTGACATCTTGGCGGGCTATATGAGACA | 900 |
| Db | 4345 | GCTGTGGACGATAGAGCTTTGGGAAGGTGCTTGTGACATCTTGGCGGGCTATATGAGACA | 4404 |
| OY | 901 | GGAATGGACGCGCGGCTCTGTGGCTTTTAAGTCAATGACGCGCGCAAAATGCCCTCACCGAG | 960 |
| Db | 4405 | GGAATGGCGCGCGGCTCTGTGGCTTTTAAGTCAATGACGCGCGCAAAATGCCCTCACCGAG | 4464 |
| OY | 961 | GACCTGGTAACTTAATCTCCCGCATCTCTCTCTGGTGGCTGATCGTGGGGGTGCTG | 1020 |
| Db | 4465 | GACCTGGTCAATCTAATCTTCTGTGCATCTCTCTCTGGCGCCCTGTGTGTGGGGGTGCTG | 4524 |
| OY | 1021 | TGCGCAGCGATATCGCTCGGACATGTGGTCAAGGGAGGGGGGCTGTGCAATGATGAAC | 1080 |
| Db | 4525 | TGTGAGGAATATCGCTCGATCGACATGTGGTCTGGGAGAGGGGGGCTGTGCAATGATGAAC | 4584 |
| OY | 1081 | CGGCTGATAGGCTTGCCTCGCGGGGGTACATATTTTCCCAAGCACTAATGTGCAAG | 1144 |
| Db | 4585 | CGGCTGATAGGCTTGCCTCGCGGGGGTAAATATATGTTCCTCCACAGCACTAATGTGCTAAG | 4644 |

| | | | |
|---|------|---|------|
| D | 3505 | TGGGAGATGTCCTTCACAGGCTCACCCATATAGATGACACTTCTTGTCACCAACCAAG | 3564 |
| Q | 61 | CAGGAGAGAGACAATTCCTCCCTACCTCGGTGGCGTACAGAGCTACTGTGTGCGTAGAGGCC | 120 |
| D | 3565 | CAGGAGAGAGACAATTCCTCCCTACCTCGGTAGCATACCAAGCAGGTGTGGCCAGAGGCT | 3624 |
| Q | 121 | CAGGCCCCCACTCCATCATATGGGATTCMAATGTGGAAAGTGTCTCATACGGCTTAACGCTACT | 180 |
| D | 3625 | CAGGCCCCCACTCCATCATATGGGATTCMAATGTGGAAAGTGTCTCATACGGCTTAACGCTACT | 3684 |
| Q | 181 | CTGGCGGGGGCCAAACAACCTTGCTGTGATATAGCTGGGAGCCGTCCAAAACGAGAGTCAACCTC | 240 |
| D | 3685 | CTGCACGGGGCCAAACAACCTTGCTGTGACGGCTGGGAGCCGTCCGAATAGAGTCAACCTC | 3744 |
| Q | 241 | ACACACCCCATTAACCAAAATTCATCAATGGCATGCAATGTCAAGCCGACCTTGAGAGTGTGACG | 300 |
| D | 3745 | ACCAACCCCATTAACCAAAATTCATCAATGGCATGCAATGTGCGCTGACCTTGAGAGTGTGACT | 3804 |
| Q | 301 | AGCACTTGGGTGTGTGTGTGGGCGGGGTCTTGACAGTCTGTGCGTGGGTATGTGTTACACAAC | 360 |
| D | 3805 | AGCACTTGGGTGTGTGTGTGGGCGGGGTCTTGACAGTCTGTGCGCGGGTATGTGTTACACAAC | 3864 |
| Q | 361 | GGCAGCGTGTGCATTGTGGGTAGATATCTTTGTCGGGAGCCGTCCAAAACGAGAGTCCCGAC | 420 |
| D | 3865 | GGCAGTGTGTGCATTGTGTGGGTAGATATCTTTGTCGGGAGCCGTCCGAATAGAGTCAACCTC | 3924 |
| Q | 421 | AGGAGAGTCTCTACCAAGAGTGTGATAGATGGAAGATCGCGGTGCACTCCCTTAC | 480 |
| D | 3925 | AGGAGAGTCTCTACCAAGAGTGTGATAGATGGAAGATCGCGCTGCACTCCCTTAC | 3984 |
| Q | 481 | ATCGAGCAGGGATATGCACTGCGCGAGCATTCACGCAAAAAGGCTCGGGTGTGCTGAC | 540 |
| D | 3985 | ATCGAGCAGGGATATGCACTGCGCGAGCAATTCACGCAAAAAGGCTCGGGTGTGCTGCA | 4044 |
| Q | 541 | ACAGGCACCAAGCAAGCGGAGCGCGCTCCCGTGTGTAGAGTCCAGATGGCGAGCCCTT | 600 |
| D | 4045 | ACAGGCACCAACAGCGGAGCGCTGCTCCCGTGTGTAGATCCAGATGGCGAGCCCTT | 4104 |
| Q | 601 | GAGACCTTCTGGGCGCAACACATGTGAACTTCATCAGCGGGATACAGTACTTAGCAGGC | 660 |
| D | 4105 | GAGACCTTCTGGGCGCAACATGTGAACTTCATCAGCGGGATACAGTACTTAGCAGGC | 4164 |
| Q | 661 | TTGTCCACTGTGCTGGGATATCCGCGCATTTGCATCATGATGGGCTTACAGCTCTGTCT | 720 |
| D | 4165 | TTATCCACTGTGCTGGGATATCCGCGCATTAGATCATGTGAGCATTCACAGCCTTATCT | 4224 |
| Q | 721 | ACTACCCCGCTACCAACCCCAATCTAACCTCCGCTTACATCTCGGGGGATATGGGTAGCC | 780 |
| D | 4225 | ACCAACCCCGCTACCAACCCCAAGTACCTCTCTGTTAATCTTGGGGGGGTGTGGTGTCT | 4284 |
| Q | 781 | GCCCAACTGTGCTCCCCCAGTGTCTTCAAGCTTTCGATGGGCGCGCATTTGCTGTGTGCG | 840 |
| D | 4285 | GCCCAACTGTGCTCCCCCAGCGCGCTTCCGCTTTCGTGTGGGCGCGGATCGCGGTGTGCG | 4344 |
| Q | 841 | GCTGTGGCAGCATAGGCTTGGGAAGTGTCTGTGCAATCTTGGCGGGCTATATGAGCA | 900 |
| D | 4345 | GCTGTGGCAGCATATAGGCTTGGGAAGTGTCTGTGCAATCTTGGCGGGCTTATATGAGCA | 4404 |
| Q | 901 | GGAATGGCAGAGCGCGCTGTGTGCTTTTAAAGTCAAGACCGCGCAAAATGCCCTCCACGAG | 960 |
| D | 4405 | GGAATGGCAGAGCGCGCTGTGTGCTTTTAAAGTCAAGACCGCGCAAAATGCCCTCCACGAG | 4464 |
| Q | 961 | GACCTGTGTTAATTTACTCCCTGCTGCATTCCTCTCTGGGCGCTGTGTGCGGGTGTG | 1020 |
| D | 4465 | GACCTGTGTCAATCTTACTTCTGTGCAATCTCTTCTCTGGGCGCTGTGTGCGGGTGTG | 4524 |
| Q | 1021 | TGCGCAGGATATCTGCGTGGCAGCGTGGTCCAGGGAGAGGGGCTGTGCAATGATGAAC | 1080 |
| D | 4525 | TGTGACGCAATATCTGCGTGCACACGTGGGTCCGGAGAGGGGGCTGTGCAATGATGAAC | 4584 |
| Q | 1081 | CGGCTGATATGCGTTGTGCTGTGCGGGGTACCATATTTTCCCGACCATATATGTGCGAG | 1140 |
| D | 4585 | CGGCTGATATGCGTTGTGCTGTGCGGGGTATCATATTTTCCCGACCATATATGTGCTGAG | 4644 |

| | | | |
|----|------|---|------|
| QY | 1141 | AGCGACGCCGAGACGTCGTACCTCAATCTCTCCGACCTTACTACACCACTGTG | 1200 |
| Db | 4645 | AGCGACGCCGAGCGCGTGTACTCAATCTCTCCGACCTTACTACCTCACTCACTCTG | 4704 |
| QY | 1201 | AAGAGGCTCCACGAGTGATTAAGAGAGCTGCTCCAGCCCTCTCGGCTGTGCTA | 1260 |
| Db | 4705 | AAAGGCTCCACGAGTGATTAAGAGAGCTGCTCCAGCCCTCTCGGCTGTGCTA | 4764 |
| QY | 1261 | AGGAGTGTGGGACTGGATATGCAAGTTTGGCTGACTTCAAGACTGGCTCAGTCC | 1320 |
| Db | 4765 | AGGAGTGTGGGACTGGATATGCAAGTTTGGCTGACTGACTTCAAGACTGGCTCAGTCC | 4824 |
| QY | 1321 | AAGCTCTGCCGCGATTAACCGGAGTCCCTTTTTCATATGCCAAGTGGGTAAAGAGG | 1380 |
| Db | 4825 | AAAGCTCTGCCGCGACTACCTGGAGTCCCTTTTTCATATGCCAAGCGGGTTAAAGAGG | 4884 |
| QY | 1381 | GTCGCGCGGGAGACGGCATCATGCAGACCACTGCTCATGTGGAGCAGATCACCGGA | 1440 |
| Db | 4885 | GTCGCGCGGGAGACGGCATCATGCAGACCACTGCTCATGTGGAGCAGATCACCGGA | 4944 |
| QY | 1441 | CATGTCAAAAACGGTTCATGAGGATCGTTGGGCTTAAGACTGTAGTAATGTGCAT | 1500 |
| Db | 4945 | CATGTCAAAAACGGTTCATGAGGATCGTGGGGCTTAAGACTGTAGTAATGTGCAT | 5004 |
| QY | 1501 | GGAACATTTCCCATCAACGCACTAACCAACCGGGCCCCCTGCACGCCCTCCGACGCCCAAC | 1560 |
| Db | 5005 | GGAACATTTCCCATCAACGCACTAACCAACCGGGCCCCCTGCACACCTTCCGACGCCCAAC | 5064 |
| QY | 1561 | TATTCACAGGGCGCTGTGGCGGGGTGGCTGTCTAGAGAGTACGTCGAGGTTTACGCGGCTGGG | 1620 |
| Db | 5065 | TATTCACAGGGCGCTGTGGCGGGGTGGCGCTGTAGAGAGTACGTCGAGGTTTACGCGGCTGGG | 5124 |
| QY | 1621 | GATTTCCACTACGTGACGAGCATGACCACTGACAAAGTAAATGCCCTGTCAGGTTCCA | 1680 |
| Db | 5125 | GATTTCCACTACGTGACGAGCATGACCACTGACAAAGTAAATGCCCATGTCAGGTTCCG | 5184 |
| QY | 1681 | GCCCCCGCAATTTCTTCAACAGAGTGAATGGGGTGGCGCTGCACAGTTACGCTCGGCGTGC | 1740 |
| Db | 5185 | GCTCCTGAATTTCTTCTCGAGGTGAGCGAGATGGCGTTGCACAGTTACGCTCGGCGTGC | 5244 |
| QY | 1741 | AAACCTCTCTACGCGGAGAGGATGCACATTCAGGTCGGGCTCAACCAATACTGTGGG | 1800 |
| Db | 5245 | AGGCTCTCTCTACGCGGAGAGGATGCATTCAGGTCGGGCTCAACCAATACTGTGGG | 5304 |
| QY | 1801 | TCGCACTTCCCATGAGAGCCGAAACCGGATGTAGCAATGCTCACTTCATGCTCACCGAC | 1860 |
| Db | 5305 | TCAGAGTACCATGAGAGCCGAAACCGGATGTAGCAATGCTCACTTCATGCTCACCGAC | 5364 |
| QY | 1861 | CCCTTCCCATCAACAGCAGAGACGGCTTAAGCGCAGGCTGGGCGCAGGGGGTCTCCCCCTCC | 1920 |
| Db | 5365 | CCCTTCCCATCAACAGCAGAGACGGCTTAAGGTTGGCGCAGGGGGTCTCCCCCTCC | 5424 |
| QY | 1921 | TTGGCAGACTCTTCAAGCTAGGCAAGTGTCTCGGCTTCTCGAAGGCGACATTAATTACC | 1980 |
| Db | 5425 | TTGGCAGACTCTTCAAGCTAGGCAAGTGTGTCTCGGCTTCTTGGAAGGCGACATTAATTACC | 5484 |
| QY | 1981 | CAAAATGACTTCCAGACGTGACCTCTATCGAGGCCAATCTCTGTGGCGGCAATGATG | 2040 |
| Db | 5485 | CACCAATGCTCTCGGAGCGTGAACCTATTCAGAGGCCAATCTCTGTGGCGGCAATGATG | 5544 |
| QY | 2041 | GGC 2043 | |
| Db | 5545 | GGC 5547 | |

RESULT 6
US-09-315-850-35
; Sequence 35, Application US/09315850
; Patent No. 6217872
; GENERAL INFORMATION:
; APPLICANT: OKAYAMA, Hiroto
; APPLICANT: FURE, Isao

APPLICANT: MORI, Chiato
 APPLICANT: TAKAMIZAWA, Akahisa
 APPLICANT: YOSHIDA, Iwao
 TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC
 TITLE OF INVENTION: CDNA AND ANTIGEN POLYPEPTIDE
 NUMBER OF SEQUENCES: 50
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Armstrong, Westerman, Hattori, McLeand &
 STREET: 1725 K St. N.W. Suite 1000
 CITY: Washington
 STATE: D.C.
 COUNTRY: U.S.A.
 ZIP: 20006
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.5 in, 1.44MB
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS, Version 5.0
 SOFTWARE: ASCII
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/315,850
 FILING DATE:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/08/904,686
 FILING DATE: 01-AUG-1997
 APPLICATION NUMBER: US 08/324,977
 FILING DATE: 18-OCT-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 2-167466
 FILING DATE: 25-JUN-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 2-230921
 FILING DATE: 31-AUG-1990
 APPLICATION NUMBER: JP 2-305605
 FILING DATE: 09-NOV-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/099,706
 FILING DATE: 30-JUL-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/769,996
 FILING DATE: 02-OCT-1991
 APPLICATION DATA:
 APPLICATION NUMBER: US 07/635,451
 FILING DATE: 28-DEC-1990
 ATTORNEY/AGENT INFORMATION:
 NAME: McLeand, Le-Nhung
 REGISTRATION NUMBER: 31,541
 REFERENCE/DOCKET NUMBER: 9007036
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 659-2930
 TELEFAX: (202) 887-0357
 INFORMATION FOR SEQ ID NO: 35:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 7863 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA from genomic RNA
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..7863
 NAME/KEY: misc_feature
 LOCATION: 1..7863
 OTHER INFORMATION: /note= "sequence = 1500 - 9362 of
 OTHER INFORMATION: SEQ ID NO: 1"
 US-09-315-850-35

Query Match 87.9%; Score 1796.6; DB 3; Length 7863;
 Best Local Similarity 92.5%; Pred. No. 0;
 Matches 1889; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

QY 1 TGGAGGGCGCTTTCACAGGCTCAACCACTGATGCCACTTCTCTGCCAACAAG 60
 DB 3505 TGGAGAGGTCTTCAAGAGCTCAACCATATAGACACTTCTTGCACAGCAAG 3564
 QY 61 CAGGACGAGAGAACTTCCCTCACTTGGTGGTACAGGCTACTGTGGCTGGGGC 120
 DB 3565 CAGGACGAGAGAACTTCCCTCACTTGGTGGTACAGGCTACTGTGGCTGGGGC 3624
 QY 121 CAGGCCCCCTCCATCATGGGATCAAAATGTGAAGTCTCTATACGGCTAAAGCTACT 180
 DB 3625 CAGGCCCCCTCCATCATGGGATCAAAATGTGAAGTCTCTATACGGCTAAAGCTACT 3684
 QY 181 CTGCGCGGCAACACCTTGTGTATAGCTGGAGCCGTCAAAACAGAGTCACTC 240
 DB 3685 CTGCAAGGGCAACACCTTGTGTATAGCTGGAGCCGTCAAAACAGAGTCACTC 3744
 QY 241 ACACACCCCTAACCAATTCTATGCGATGATGTACGCGACTGGAAGTCTGACG 300
 DB 3745 ACCCACCCCTAACCAATTCTATGCGATGATGTACGCGACTGGAAGTCTGACCT 3804
 QY 301 AGCACCCTGGGTGCTGGTGGGCGGGGCTTGACGCTTGAGCTTGGCTTATGTTCAACA 360
 DB 3805 AGCACCCTGGGTGCTGGTGGGCGGGGCTTGAGCTTGGCTTATGTTCAACA 3864
 QY 361 GGCAGCGTGTCTATGTGGGTAGATCATCTGTCCGGCGGCGGCTATGTTCCGAC 420
 DB 3865 GGCAGCGTGTCTATGTGGGTAGATCATCTGTCCGGCGGCGGCTATGTTCCGAC 3924
 QY 421 AGGGAATCTCTTACCAAGAGTTGATGATGAGAAAGTGGCGTCCGACTCCCTTAC 480
 DB 3925 AGGGAATCTCTTACCAAGAGTTGATGATGAGAAAGTGGCGTCCGACTCCCTTAC 3984
 QY 481 ATCGACAGGGAATGAGCTGCGGAGCTCAAGCAAAAGCGCTGGGTGTCAG 540
 DB 3985 ATCGACAGGGAATGAGCTGCGGAGCTCAAGCAAAAGCGCTGGGTGTCAG 4044
 QY 541 ACAGCCACCAACCAAGCGAGGCGCTGCTCCGTGTGGAGTCCAGTGGGAGCCCTT 600
 DB 4045 ACAGCCACCAACCAAGCGAGGCGCTGCTCCGTGTGGAGTCCAGTGGGAGCCCTT 4104
 QY 601 GAGACCTTGGGGCGAAACATGTGAACTTCATCAGCGGATACAGTACTTAGCAGGC 660
 DB 4105 GAGACCTTGGGGCGAAACATGTGAACTTCATCAGCGGATACAGTACTTAGCAGGC 4164
 QY 661 TTGTCCACTCTGCGGGAATCCCGGATGTGATAGATGATGCGCTTCACAGCTCTGTC 720
 DB 4165 TTATCCACTCTGCGGGAATCCCGGATGTGATAGATGATGCGCTTCATC 4224
 QY 721 ACTAGCCGCTCACACCAATCTACCTCTGTTAAATCTGCGGGGATGGGTAGCC 780
 DB 4225 ACCAGCCGCTCACACCAATCTACCTCTGTTAAATCTTGGGGGGTGGGTGCT 4284
 QY 781 GCCCACTGCTCCCCAGTCTGCTTCAAGTTTGTAGAGCGCGCATTGTGTGTCG 840
 DB 4285 GCCCACTGCTCCCCAGTCTGCTTCAAGTTTGTAGAGCGCGCATTGTGTGTCG 4344
 QY 841 GCTGTGGAGATAGCTTGGGAAGTCTTGGAGATCTTGGCGGCTATGAGAGA 900
 DB 4345 GCTGTGGAGATAGCTTGGGAAGTCTTGGAGATCTTGGCGGCTATGAGAGA 4404
 QY 901 GGAGTGGCAGGCGGCTGCTGCTTAAAGTCAATGAGCGGGAATGCGCTCCACCGAG 960
 DB 4405 GGAGTGGCAGGCGGCTGCTGCTTAAAGTCAATGAGCGGGAATGCGCTCCACCGAG 4464
 QY 961 GACTGTGTAATTAATCTCTTCCATCTCTTCTGCTGCTGCTGCTGCTGCTGCTG 1020
 DB 4465 GACTGTGTAATTAATCTCTTCCATCTCTTCTGCTGCTGCTGCTGCTGCTGCTG 4524
 QY 1021 TGGCAGAGATCTGCGTGGAGCGTGGGTCACAGGGGAGGGGCTGTGAGTGAATG 1080
 DB 4525 TGGCAGAGATCTGCGTGGAGCGTGGGTCACAGGGGAGGGGCTGTGAGTGAATG 4584
 QY 1081 CGGCTGATAGCGTTCGCGCGGGGTAAACATGTTTCCCAACGACATATGTGCCAGAG 1140

```

Db      4585  CCGCTGATAGCGCTCGCGGGGTATCATGTTTCCCGACACGACTAGTGCTGAG 4644
Oy      1141  AGCGAGCGCGAGCAGTGTCACTAGATCTCTCCGACCTTACTATCAACCACTGTTG 1200
Db      4645  AGCGAGCGCGAGCGCGGTCTTCTAGATCTCTCCAGCTTACATCACTAGTGCTG 4704
Oy      1201  AAGAGGCTTCACAGTGTATTAACAGAGACTCTCCAGCCCTGCTCGGGTCTGAGCTA 1260
Db      4705  AAAAGGCTTCACAGTGTATTAATGAAGACTCTCCACACCGTGTCCGGCTCGAGCTA 4764
Oy      1261  AGGAGTGTGGAGCTGATATGACAGTGTGGCTGATTAAGACCTGGCTCCAGTCC 1320
Db      4765  AGGGATGTTGGAGCTGATATGACAGGCTGTGACTGACTTAAAGCTGGCTCCAGTCC 4824
Oy      1321  AAGCTCTGCGCGAGTTACCGGAGTCCCTTTTCTCATGCGCAAGTGGGTACAAGGG 1380
Db      4825  AAGCTCTGCGCGAGTACCTGAGTCCCTTTTCTCGTGCGCAAGCGGGTACAAGGA 4884
Oy      1381  GTCTGGCGGGAGAGCGGCATCATGCAAAACCACTGCCATGTGGAGCAGATCACCGGA 4944
Db      4885  GTCTGGCGGGAGAGCGGCATCATGCAAAACCACTGCCATGTGGAGCAGATCACCGGA 4944
Oy      1441  CATGTCAAAAAGGTTTCCATAGAGATCGTTGGGCTTAAGACCTGTATGAATGTGGAT 1500
Db      4945  CATGTCAAAAAGGTTTCCATAGAGATCGTTGGGCTTAAGACCTGTGAAGACCAACG 5004
Oy      1501  GGAACATTTCCCATCAACGATACACCAAGGAGCCCTGACCGCCCTCCCGACGCGCAAC 1560
Db      5005  GGAACATTTCCCATCAACGATACACCAAGGAGCCCTGACACCTCTCCACGCGCAAC 5064
Oy      1561  TATTCAGAGCGCTGTGGCGGTGTGCTGTAGAGATCATGTGAGGTTACGCGGGTGGG 1620
Db      5065  TATTCAGAGCGCTGTGGCGGTGTGCTGTAGAGATCATGTGAGGTTACGCGGGTGGG 5124
Oy      1621  GATTTCCACTAGTACGAGCATGACCACTGACCAAGTAAATGCCCGTCCAGGTTCCA 1680
Db      5125  GATTTCCACTAGTACGAGCATGACCACTGACCAAGTAAATGCCCGTCCAGGTTCCG 5184
Oy      1681  GCCCCGGAATTCCTCAAGAGTGTGAGGAGTGGGCTGACAGAGTACGCTCCGGCTGTC 1740
Db      5185  GCTCTGAATTTCTTCGAGAGTGTGACGAGTGGCGTTGACAGAGTACGCTCCGGCTGTC 5244
Oy      1741  AAACCTCTCTCAAGGAGAGGTCACTTCAAGTCCGAGTCCAGTCAACCAATACCTGTTGG 1800
Db      5245  AAGCCCTCTCTCAAGGAGAGGTACATTCAGAGTGGGCTCAACCAATACCTGTTGG 5304
Oy      1801  TGGAGCTCCCATGCGAGCCCGAATGTAAGAGTGTCTCACTTCCATGCTCAACCGAC 1860
Db      5305  TCACAGCTTCAATGCGAGCCCGAATGTAAGAGTGTCTCACTTCCATGCTCAACCGAC 5364
Oy      1861  CCTCCCATCAACAGAGAGAGGCTGAGCGAGGCTGAGCGAGGAGTCTCCCGCTCC 1920
Db      5365  CCTCCCATCAACAGAGAGAGGCTGAGCGAGGCTGAGCGAGGAGTCTCCCGCTCC 5424
Oy      1921  TTGGCAGGCTTTCACTAGCAAGTGTGTGCGCTTCTCTGAAAGCGCATACATTACC 1980
Db      5425  TTGGCAGGCTTTCACTAGCAAGTGTGTGCGCTTCTCTGAAAGCGCATACATTACC 5484
Oy      1981  CAAATGATTTCCCAAGAGCTGACCTCATGAGGCGCAACCTCTGTGGGGGATAGAG 2040
Db      5485  CACCATGTCTTCCGAGCGCTGACCTCATGAGGCGCAACCTCTGTGGGGGATAGAG 5544
Oy      2041  GGC 2043
Db      5545  GGC 5547

```

```

RESULT 7
US-08-324-977-31
; Sequence 31, Application US/08324977
; Patent No. 5747339
; GENERAL INFORMATION:

```

```

; APPLICANT: OKAYAMA, Hiroto
; APPLICANT: FUKE, Isao
; APPLICANT: MORI, Chisato
; APPLICANT: TAKAMIZAWA, Akahisa
; APPLICANT: YOSHIDA, Iwao
; TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC
; TITLE OF INVENTION: cDNA AND ANTIGEN POLYPEPTIDE
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESS: Armstrong, Westerman, Hatlori, McLealand &
; ADDRESSEE: Naughton
; STREET: 1725 K St. N.W. Suite 1000
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 in, 1.44MB
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS, Version 5.0
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/324,977
; FILING DATE: 18-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-167466
; FILING DATE: 25-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-230921
; FILING DATE: 31-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-305605
; FILING DATE: 09-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/099,706
; FILING DATE: 30-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/769,996
; FILING DATE: 02-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/635,451
; FILING DATE: 28-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Stevens-Smith, Theresa M.
; REGISTRATION NUMBER: 36,281
; REFERENCE/DOCKET NUMBER: 900703D
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 659-2930
; TELEFAX: (202) 887-0357
; TELEX: 440142
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7917 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA from genomic RNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..7862
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..7916
; OTHER INFORMATION: /note= "sequence = 1500 - 9416 of
; OTHER INFORMATION: SEQ ID NO: 1"
; US-08-324-977-31

```

```

Query Match      87.9%; Score 1796.6; DB 1; Length 7917;
Best Local Similarity 92.5%; Pred. No. 0;
Matches 1889; Conservative 0; Mismatches 154; Indels 0; Gaps 0;
Oy      1  TTGGAGGCGCTTTACAGCGCTTACCGAGTGGATGCGCACTTCTGTTCCCAACAAG 60

```

Db 3505 TGGGAGAGTGTCTTCAAGGCTTCAACCATATAGATGACAATTCTTGTCCAGACCAAG 3564
 QY 61 CAGGAGAGAGCAACTTCCCTTACCTGAGGCGTACAGGCTTACTGTGCGCTAGAGGC 120
 Db 3565 CAGGAGAGAGCAACTTCCCTTACCTGAGGCGTACAGGCTTACTGTGCGCTAGAGGC 3624
 QY 121 CAGGCCCCACCTTCATCATGAGATCAAAATGTGGAAGTGTCTCATACAGGCTTAAAGCTTACT 180
 Db 3625 CAGGCCCCACCTTCATCATGAGATCAAAATGTGGAAGTGTCTCATACAGGCTTAAAGCTTACT 3684
 QY 181 CTGCGCGGCGCAACACCTTCTGTATAGGCTGGAGCGCTCCAAACAGAGTCAACCTTC 240
 Db 3685 CTGCAACGGGCAACACCTTCTGTATAGGCTGGAGCGCTCCAAACAGAGTCAACCTTC 3744
 QY 241 ACACACCCCATTAACAAATTCATCATGAGATGATGATGACGCGACCTGGAGGTCGTCAG 300
 Db 3745 ACCACCCCATTAACAAATTCATCATGAGATGATGATGACGCGACCTGGAGGTCGTCAG 3804
 QY 301 AGCAGCTGGGCTGTGTGGCGGGGTCTTGCAGCTCTGGCTGCTGATTTGCTTGAACACA 360
 Db 3805 AGCAGCTGGGCTGTGTGGCGGGGTCTTGCAGCTCTGGCTGCTGATTTGCTTGAACACA 3864
 QY 361 GCGAGCGTGTCTATGTGGGTAGGATCATCTTGTCCGGCGCGCTATTTGTTCCGAC 420
 Db 3865 GCGAGTGTGTCTATGTGGGTAGGATCATCTTGTCCGGCGCGCTATTTGTTCCGAC 3924
 QY 421 AGGGAAGTCTCTACAGGAGTTTCAATGAGATGGAAGTGGCGGTGACCTCCCTTAC 480
 Db 3925 AGGGAAGTCTCTACAGGAGTTTCAATGAGATGGAAGTGGCGGTGACCTCCCTTAC 3984
 QY 481 ATCGACAGGGAATGAGCTCGCCGAGCAGTTCAAGCAAAAGCGCTCGGTTGCTGAC 540
 Db 3985 ATCGACAGGGAATGAGCTCGCCGAGCAGTTCAAGCAAAAGCGCTCGGTTGCTGAC 4044
 QY 541 ACAGCCCAACAGCAAGCGAGCGCTGCTCCGCTGTGAGAGTCCAAAGTGGCGAGCCTT 600
 Db 4045 ACAGCCCAACAGCAAGCGAGCGCTGCTCCGCTGTGAGAGTCCAAAGTGGCGAGCCTT 4104
 QY 601 GAGACCTTCTGGGCAAAACATGTGGAATCTTCAATGAGGGAATACAGTACTTACAGAGC 660
 Db 4105 GAGACCTTCTGGGCAAAACATGTGGAATCTTCAATGAGGGAATACAGTACTTACAGAGC 4164
 QY 661 TTGTCACTCTGCTGGGAATCCCGGATTTGATCATGATGAGCGCTTACAGCCTCTATC 720
 Db 4165 TTATCACTCTGCTGGGAATCCCGGATTTGATCATGATGAGCGCTTACAGCCTCTATC 4224
 QY 721 ACTAGCCGCTCAACCAATCTACCTCTCTGCTTAAATCTCTGAGGAGATGAGTACGC 780
 Db 4225 ACTAGCCGCTCAACCAATCTACCTCTCTGCTTAAATCTCTGAGGAGATGAGTACGC 4284
 QY 781 GCCCAATCTGCTCCCGCAATGCTGTGATGCTTTCGTAGGCGCGGCAATTTGCTGTGG 840
 Db 4285 GCCCAATCTGCTCCCGCAATGCTGTGATGCTTTCGTAGGCGCGGCAATTTGCTGTGG 4344
 QY 841 GCTGTGGAGATAGAGCTTGGAAAGTGTGTGACATCTTGGCGGCTATGAGAGA 900
 Db 4345 GCTGTGGAGATAGAGCTTGGAAAGTGTGTGACATCTTGGCGGCTATGAGAGA 4404
 QY 901 GAGAGTGGAGGCGCTGTGGCTTTTAAAGTCAATGAGGCGGAAATGCTTCCACCGAG 960
 Db 4405 GAGAGTGGAGGCGCTGTGGCTTTTAAAGTCAATGAGGCGGAAATGCTTCCACCGAG 4464
 QY 961 GACCTGTGTAATCTTACTCCCTGCAATCTCTCTCTGAGGCGCTGTGCTGGGGTCTGG 1020
 Db 4465 GACCTGTGTAATCTTACTCCCTGCAATCTCTCTCTGAGGCGCTGTGCTGGGGTCTGG 4524
 QY 1021 TGGCAGCGATACTGCTGAGCAGTGGTCCAGGGAGGGGCTGTGAGTGAATGAAC 1080
 Db 4525 TGTGAGCAATACTGCTGAGCAGTGGTCCAGGGAGGGGCTGTGAGTGAATGAAC 4584
 QY 1081 CGGCTGATAGCTTGGCTCGCGGGGTAAACATGTTTTCCCAACGCACTATGTGCCAGAG 1140
 Db 4585 CGGCTGATAGCTTGGCTCGCGGGGTAAACATGTTTTCCCAACGCACTATGTGCCAGAG 4644

QY 1141 AGCGAGCGCGAGCAGTGTCACTCAGATCTCTCCGACTTACTATCAACCAACTGTG 1200
 Db 4645 AGCGAGCGCGAGCAGCGGTGTTACTAGATCTCTCCAGCTTACATCACTAGCTGCG 4704
 QY 1201 AAGAGCTTCAACCAATGATTAACAGAGACTGCTCAAGCCTGTCTCCGGCTGTGGCTA 1260
 Db 4705 AAAAGCTTCAACCAATGATTAATGAAAGCTCTCCACACCGTGTCTCCGGCTGTGGCTA 4764
 QY 1261 AAGGATGTTTGGAGCTGGAATAGCAAGTGTGGGCTTGAACCTGGGCTCCAGTCC 1320
 Db 4765 AAGGATGTTTGGAGCTGGAATAGCAAGTGTGGGCTTGAACCTGGGCTCCAGTCC 4824
 QY 1321 AAGCTCTGCGCGCAGTACCGGAGTCCCTTTTCTCATGCGCAAGTGGTAAAGGG 1380
 Db 4825 AAGCTCTGCGCGCAGTACCGGAGTCCCTTTTCTCATGCGCAAGTGGTAAAGGG 4884
 QY 1381 GTCTGGCGGGAGAGCGGATCATGAGACCACTGCTCATGTGAGCAGATCAACCGGA 1440
 Db 4885 GTCTGGCGGGAGAGCGGATCATGAGACCACTGCTCATGTGAGCAGATCAACCGGA 4944
 QY 1441 CATGTCAAAAAGGTTCCATGAGATGCTGTGGGCTTAAAGCCTGTATGATCAATGGCAT 1500
 Db 4945 CATGTCAAAAAGGTTCCATGAGATGCTGTGGGCTTAAAGCCTGTATGATCAATGGCAT 5004
 QY 1501 GGAACATTTCCCATCAACGATCAACAGAGGCGCTGACGCGCTCCAGAGCGCAAA 1560
 Db 5005 GGAACATTTCCCATCAACGATCAACAGAGGCGCTGACGCGCTCCAGAGCGCAAA 5064
 QY 1561 TATTCAGAGCGCTGTGGCGGGTGTCTGTAGAGATGATGAGAGTTACCGGGGTGGG 1620
 Db 5065 TATTCAGAGCGCTGTGGCGGGTGTCTGTAGAGATGATGAGAGTTACCGGGGTGGG 5124
 QY 1621 GATTTTCACTAGTGAAGAGTGAACACTGACAACTGAAATGCCCCGCTCAGGTTCCA 1680
 Db 5125 GATTTTCACTAGTGAAGAGTGAACACTGACAACTGAAATGCCCCGCTCAGGTTCCG 5184
 QY 1681 GCCCGCAATTTCTTCAAGAAATGATGATGGGCGGCTGACAGGTAAGCTCCGCGTGC 1740
 Db 5185 GCTCTGAATTTTCTTCCGAGGATGGAAGGATGCGGCTGTGACAGGTAAGCTCCGCGTGC 5244
 QY 1741 AAACCTCTCTTACGGAAGAGGTCAATTTCCAGTCCGAGCTTCAACCAATCTCTGTGG 1800
 Db 5245 AAGCGCTCTCTTACGGAAGAGGTCAATTTCCAGTCCGAGCTTCAACCAATCTCTGTGG 5304
 QY 1801 TGCAGCTTCCATGAGAGCGGAAACCGGATGTAGAGATGATCTTCCATGCTCACCGAC 1860
 Db 5305 TCACAGCTTCCATGAGAGCGGAAACCGGATGTAGAGATGATCTTCCATGCTCACCGAC 5364
 QY 1861 CCTCCCATCATGAGAGAGAGCGGCTAAGCGCAGGCTGAGCCAGGGGCTCTCCCGCTCC 1920
 Db 5365 CCTCCCATCATGAGAGAGAGCGGCTAAGCGGCTGAGGCTTGGCCAGGGGCTCTCCCGCTCC 5424
 QY 1921 TTGGCAGCTCTTCACTAGCAGTGTCTGCGCTTCTCGAAGCGCATATATAC 1980
 Db 5425 TTGGCAGCTCTTCACTAGCAGTGTCTGCGCTTCTCGAAGCGCATATATAC 5484
 QY 1981 CAAATGACTTCCAGAGGCTGACTCATCGAGGCAACCTCTGTGGCGGATAGAGATG 2040
 Db 5485 CACCATGTCTTCCGAGAGCTGACTCATCGAGGCAACCTCTGTGGCGGAGAGATG 5544
 QY 2041 GGC 2043
 Db 5545 GGC 5547

RESULT 8
 US-08-384-616-31
 ; Sequence 31, Application US/08384616
 ; Patent No. 5847101
 ; GENERAL INFORMATION:
 ; APPLICANT: OKAYAMA, Hiroto
 ; APPLICANT: FUKE, Iseao


```

QY 1201 AAGAGCTCCACGATGATTAAGAGACTGCTCCAGCCCTGCTCGGCTGTGAGCTA 1260
DB 4705 AAAAGCTCCACGATGATTAAGAGACTGCTCCAGCCCTGCTCGGCTGTGAGCTA 4764
QY 1261 AAGGATGTTGGAGATGATATGACAGATTGGCTGACTTCAAGACCTGGCTCCAGTCC 1320
DB 4765 AAGGATGTTGGAGATGATATGACAGATTGGCTGACTTCAAGACCTGGCTCCAGTCC 4824
QY 1321 AAGCTCTGCGCGGATTAACCGGAGTCCCTTTTTCATGCAAGAGTGGGTCAAGAGGG 1380
DB 4825 AAGCTCTGCGCGGATTAACCGGAGTCCCTTTTTCATGCAAGAGTGGGTCAAGAGGG 4884
QY 1381 GTCTGGCGGAGAGACGCGATCATGACAGCACTGTCTCATGTGAGACAGATCACCGGA 1440
DB 4885 GTCTGGCGGAGAGACGCGATCATGACAGCACTGTCTCATGTGAGACAGATCACCGGA 4944
QY 1441 CATGTCAAAAACGGTTCCATGAGAGATGTTGGGCTTAAGACTGTAGTAAATCATGTGGCAT 1500
DB 4945 CATGTCAAAAACGGTTCCATGAGAGATGTTGGGCTTAAGACTGTAGTAAATCATGTGGCAT 5004
QY 1501 GGAACATTTCCCATCAACGATACACAGAGGCGCCCTGACAGCGCCCTCCAGGCGCAAC 1560
DB 5005 GGAACATTTCCCATCAACGATACACAGAGGCGCCCTGACAGCGCCCTCCAGGCGCAAC 5064
QY 1561 TATTCAGGCGCGCTGTGCGGCTGTGCTGTGAGAGATGAGAGGTTACGCGGCTGGG 1620
DB 5065 TATTCAGGCGCGCTGTGCGGCTGTGCTGTGAGAGATGAGAGGTTACGCGGCTGGG 5124
QY 1621 GATTTCCACTACGTGACAGACATGACCACTGACAGATGTAATGCGCGTGCAGGTTCCA 1680
DB 5125 GATTTCCACTACGTGACAGACATGACCACTGACAGATGTAATGCGCGTGCAGGTTCCA 5184
QY 1681 GCGCCCGAATTTCTTCAGAGATGATGGGGTGGCGGCTGACAGATGACCTCCGGGCTGC 1740
DB 5185 GCGCCCGAATTTCTTCAGAGATGATGGGGTGGCGGCTGACAGATGACCTCCGGGCTGC 5244
QY 1741 AAACCTCTCTCTACGAGAGAGGTCACTTTCAGAGTGGGCTCAACCAATCTGTTGGG 1800
DB 5245 AAGGCTCTCTCTACGAGAGAGGTCACTTTCAGAGTGGGCTCAACCAATCTGTTGGG 5304
QY 1801 TCGCAGCTCCCATGCGAGCCCGAACCAGATGTAAGACATGCTCACTTCATGCTACCGAC 1860
DB 5305 TCGCAGCTCCCATGCGAGCCCGAACCAGATGTAAGACATGCTCACTTCATGCTACCGAC 5364
QY 1861 CCCTCCCATATACAGCAGACAGCGGTAACCGGAGCTGGCCGAGGGGTCTCCCTCCCTCC 1920
DB 5365 CCCTCCCATATACAGCAGACAGCGGTAACCGGAGCTGGCCGAGGGGTCTCCCTCCCTCC 5424
QY 1921 TTGGCAGCTCTTACGCTAGCCAGTTGTCTGGCGCTTCTTCGAAGGCGCATCATTTACC 1980
DB 5425 TTGGCAGCTCTTACGCTAGCCAGTTGTCTGGCGCTTCTTCGAAGGCGCATCATTTACC 5484
QY 1981 CAAAATGACTTCCAGACGCTGACCTCATGAGGCGCAACCTCTGTGGCGGCGATGAGATG 2040
DB 5485 CACCATGTCTCTCCGAGCGCTGACCTCATGAGGCGCAACCTCTGTGGCGGCGGAGATG 5544
QY 2041 GGC 2043
DB 5545 GGC 5547

```

RESULT 9

```

US-08-904-686A-31
; Sequence 31, Application US/08904686A
; Patent No. 598130
; GENERAL INFORMATION:
; APPLICANT: OKAYAMA, Hiroto
; APPLICANT: FUKE, Iisao
; APPLICANT: MORI, Chisato
; APPLICANT: TAKAMIZAWA, Akahisa
; APPLICANT: YOSHIDA, Iwao
; TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC
; TITLE OF INVENTION: CDNA AND ANTIGEN POLYPEPTIDE

```

```

; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Armstrong, Westerman, Hattori, Mcleland &
; ADDRESSEE: Naughton
; STREET: 1725 K St. N.W. Suite 1000
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 in, 1.44MB
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS, Version 5.0
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/904,686A
; FILING DATE: 01-AUG-1997
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/324,977
; FILING DATE: 18-OCT-1994
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: JP 2-167466
; FILING DATE: 25-JUN-1990
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: JP 2-230921
; FILING DATE: 31-AUG-1990
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: JP 2-305605
; FILING DATE: 09-NOV-1990
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/099,706
; FILING DATE: 30-JUL-1993
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/769,996
; FILING DATE: 02-OCT-1991
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/635,451
; FILING DATE: 28-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Mcleland, Le-Nhung
; REGISTRATION NUMBER: 31,541
; REFERENCE/DOCKET NUMBER: 900703G
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 659-2930
; TELEFAX: (202) 687-0357
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7917 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULAR TYPE: CDNA from genomic RNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..7862
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1..7916
; OTHER INFORMATION: /note="sequence = 1500 - 9416 of
; OTHER INFORMATION: SEQ ID NO: 1"
; US-08-904-686A-31

```

```

Query Match      87.9%; Score 1796.6; DB 2; Length 7917;
Best Local Similarity 92.5%; Pred. No. 0;
Matches 1889; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

```

```

QY 1 TGGAGGGGCGCTTTCACAGGCTCCACCGAGTGATGCCACTTCTGTCGAACAAG 60
DB 3505 TGGAGAGGTGTCTTACAGGCTTCAACCATATAGATGACACTTTTGTCCAGACCAAG 3564
QY 61 CAGGAGAGAGACACTTCCCTTACCTGTGTGGCGTACAGGCTACTGTGTGCGCTAAGGCC 120
DB 3565 CAGGAGAGAGACACTTCCCTTACCTGTGTGGCGTACAGGCTACTGTGTGCGCGCAGGGCT 3624

```


QY 121 CAGGCCCACTTCATCATGGATCAAAATGTGAAGTGTCTATACGGCTAAAGCTACT 180
 Db 3625 CAGGCCCACTTCATCATGGATCAAAATGTGAAGTGTCTATACGGCTAAAGCTACTAG 3684
 QY 181 CTGCGCGGCGCAACACCTTGTCTGTATAGGCTGGAGCCGCTCCAAAAGAGTCAACCTTC 240
 Db 3685 CTGCAACGGGCGCAACACCTTGTCTGTATAGGCTGGAGCCGCTCCAAAAGAGTCAACCTTC 3744
 QY 241 ACACACCCCATTAACAAATTCATCATGGATCATGTCAAGCCGACCTGGAGGTCGTCAAG 300
 Db 3745 ACCCAACCCCATTAACAAATTCATCATGGATCATGTCAAGCCGACCTGGAGGTCGTCACT 3804
 QY 301 AGCACCTGGGTCCTGGTGGGCGGGGCTCTTGACGCTCTGCGCTGCGTATTGCTTGAACA 360
 Db 3805 AACACCTGGGTCCTGGTGGGCGGGGCTCTTGACGCTCTGCGCTGCGTATTGCTTGAACA 3864
 QY 361 GGCACGCTGTGTATTTGTGTGTAGATCATCTTGTCCGGGCGGCGGCTATTGTCTCCAG 420
 Db 3865 GGCAGTGTGTATTTGTGTGTAGATCATCTTGTCCGGGCGGCGGCTATTGTCTCCAG 3924
 QY 421 AGGGAAGTCTCTACAGAGATTCATGATGAGATGGAAGATGGCGCGTCCGACCTCCCTTAC 480
 Db 3925 AGGGAAGTCTCTACAGAGATTCATGATGAGATGGAAGATGGCGCGTCCGACCTCCCTTAC 3984
 QY 481 ATCGACAGGGAATGACGCTCGCCGACAGTTCAGCAAAAAGCGCTCGGTTGCTGCAG 540
 Db 3985 ATCGACAGGGAATGACGCTCGCCGACAGTTCAGCAAAAAGCGCTCGGTTGCTGCAG 4044
 QY 541 AAGCCACCAACAGACAGCGAGCGCTGTCTCCCGTGTGTAGATCCAGTGTGGAGCCCTT 600
 Db 4045 AAGCCACCAACAGACAGCGAGCGCTGTCTCCCGTGTGTAGATCCAGTGTGGAGCCCTT 4104
 QY 601 GAGACCTTCTGGGCGGAAACACATGTGTGAATTCATCATGGCGGATACAGTACTTGAAGAG 660
 Db 4105 GAGACCTTCTGGGCGGAAACACATGTGTGAATTCATCATGGCGGATACAGTACTTGAAGAG 4164
 QY 661 TTGTCACTCTGCTGGGAATCCCGGATGTGATCACTGATGGCGTTCAAGCTCTGTCT 720
 Db 4165 TTATCACTCTGCTGGGAATCCCGGATGTGATCACTGATGGCGTTCAAGCTCTGTCT 4224
 QY 721 ACTAGCCGCTCAACCAACCAATTCATCTCTCTCTTAAATCTCTGGGGGAGATGGGTAGCC 780
 Db 4225 ACCAGCCGCTCAACCAACCAATTCATCTCTCTCTTAAATCTCTGGGGGAGATGGGTAGCC 4284
 QY 781 GCCCAACTGCTCCCCGAGTGTCTGAGCTTTGTGTAGGCGCGGCAATTCGTGGTGG 840
 Db 4285 GCCCAACTGCTCCCCGAGTGTCTGAGCTTTGTGTAGGCGCGGCAATTCGTGGTGG 4344
 QY 841 GCTGTGTGAGCATAGGCTTTGGGAAGTGTGTGTGACATCTTGGCGGGCTATAGAGCA 900
 Db 4345 GCTGTGTGAGCATAGGCTTTGGGAAGTGTGTGTGACATCTTGGCGGGCTATAGAGCA 4404
 QY 901 GGAATGTGAGGCGCGCTGT 960
 Db 4405 GGAATGTGAGGCGCGCTGT 4464
 QY 961 GACTGTGTAACTTACTCTCTGCTGATCTCTCTCTGATGCTGTGTGTGTGTGTGTGTGT 1020
 Db 4465 GACTGTGTAACTTACTCTCTGCTGATCTCTCTCTGATGCTGTGTGTGTGTGTGTGTGT 4524
 QY 1021 TGCAGCAGATATCTGCTGCGGCACTGTGGTTCAGGGGAGGGGGCTGTGAGTGTAGTAA 1080
 Db 4525 TGTGACAGCAATATCTGCTGCGGCACTGTGGTTCAGGGGAGGGGGCTGTGAGTGTAGTAA 4584
 QY 1081 CGGCTGATATGCTTCCGCTGCGGGGTAACCATTTTCCCGCAAGCATATGTGCCAGAG 1140
 Db 4585 CGGCTGATATGCTTCCGCTGCGGGGTAACCATTTTCCCGCAAGCATATGTGCCAGAG 4644
 QY 1141 AGGAGCGCGGAGCAGTGTCACTGATCTCTCTCGAAGCTTATCTATCAACCAAGCTGTG 1200
 Db 4645 AGGAGCGCGGAGCAGTGTCACTGATCTCTCTCGAAGCTTATCTATCACTACTGCTGTG 4704

QY 1201 AAGAGCTCCACAGTGTATTACAGAGATCTGCTCCAGCCCTGTCCGGCTCGTGGCTA 1260
 Db 4705 AAAAGGCTCCACAGTGTATTACAGAGATCTGCTCCAGCCCTGTCCGGCTCGTGGCTA 4764
 QY 1261 AGGAGTGTGGGACTGTGATATGCAAGTTTGTGTGCTGCTCAAGACCTGTGCTCAAGTCC 1320
 Db 4765 AGGAGTGTGGGACTGTGATATGCAAGTTTGTGTGCTGCTCAAGACCTGTGCTCAAGTCC 4824
 QY 1321 AAGCTCTGCGCGATTTACCGGGAGTCCCTTTTCTGATGCCAACGTGGTACAGGGG 1380
 Db 4825 AAGCTCTGCGCGATTTACCGGGAGTCCCTTTTCTGATGCCAACGTGGTACAGGGG 4884
 QY 1381 GTCTGGCGGGAGAGACGATATCATGCAACACCTGTCTATGTGAGCAACATACCGGGA 1440
 Db 4885 GTCTGGCGGGAGAGACGATATCATGCAACACCTGTCTATGTGAGCAACATACCGGGA 4944
 QY 1441 CATGTCAAAAACGGTTCATGAGATCGTGTGGGCTTAAGACCTGTATGTAATGTGAT 1500
 Db 4945 CATGTCAAAAACGGTTCATGAGATCGTGTGGGCTTAAGACCTGTATGTAATGTGAT 5004
 QY 1501 GGAACATTCCCATTCACAGATACACAGCGGCGCTTGACGCGCTCCAGCGCAAA 1560
 Db 5005 GGAACATTCCCATTCACAGATACACAGCGGCGCTTGACGCGCTCCAGCGCAAA 5064
 QY 1561 TATTCAGGGGCTGTGGCGGCTGTCTGTGAGAGATACGTGAGTTAACCGCGGTGGG 1620
 Db 5065 TATTCAGGGGCTGTGGCGGCTGTCTGTGAGAGATACGTGAGTTAACCGCGGTGGG 5124
 QY 1621 GATTTCCACTACGTGACAGCATGACACTGACCAACGTAAATGCGCGCTGCAGGTTTCA 1680
 Db 5125 GATTTCCACTACGTGACAGCATGACACTGACCAACGTAAATGCGCGCTGCAGGTTTCA 5184
 QY 1681 GCCCCGGAATCTTTCACAGAAATGTGATGGGTGTGGCTGCAAGGTACGTTCCGGCTGC 1740
 Db 5185 GCTCTGAATCTTCTCGAGAGTGGACGAGGTGCGGTTGCAAGGTACGTTCCGGCTGC 5244
 QY 1741 AAACCTCTCTACCGGAGAGAGTCACTTCCAGGTGTGGGCTTCAACCAATTCCTGTTGG 1800
 Db 5245 AAGCTCTCTCTACCGGAGAGAGTCACTTCCAGGTGTGGGCTTCAACCAATTCCTGTTGG 5304
 QY 1801 TCGAGCTCCCATGCGAGCCCGAAGCGGATGTAGAGTGTCACTTCCATGCTCACCGAC 1860
 Db 5305 TCGAGCTCCCATGCGAGCCCGAAGCGGATGTAGAGTGTCACTTCCATGCTCACCGAC 5364
 QY 1861 CCTTCCCATTCACAGAGAGAGCGGTAAGCGCAAGCTGCGCAAGGGGTCTTCCCTCTCC 1920
 Db 5365 CCTTCCCATTCACAGAGAGAGCGGTAAGCGGTAAGCTGCGCAAGGGGTCTTCCCTCTCC 5424
 QY 1921 TTGGCCAGCTCTTACGCTAGGCAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1980
 Db 5425 TTGGCCAGCTCTTACGCTAGGCAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5484
 QY 1981 CAAAATGACTTCCAGAGCGTGAAGCTCATCGAGGCAACCTCTGTGCGGATGAGATG 2040
 Db 5485 CAACATGTCTTCCGAGCGCTGACCTCATCGAGGCAACCTCTGTGCGGATGAGATG 5544
 QY 2041 GGC 2043
 Db 5545 GGC 5547

RESULT 10
 US-09-315-850-31
 ; Sequence 31, Application US/09315850
 ; Patent No. 6217872
 ; GENERAL INFORMATION:
 ; APPLICANT: OKAYAMA, Hiroto
 ; APPLICANT: FUKE, Isao
 ; APPLICANT: MORI, Chisato
 ; APPLICANT: TAKAMIZAWA, Akahisa
 ; APPLICANT: YOSHIDA, Iwao
 ; TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC
 ; ;
 ; TITLE OF INVENTION: CDNA AND ANTIGEN POLYPEPTIDE

NUMBER OF SEQUENCES: 50
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Armstrong, Westerman, Hattori, Mcleland &
 ADDRESSEE: Naughton
 STREET: 1725 K St. N.W. Suite 1000
 CITY: Washington
 STATE: D.C.
 COUNTRY: U.S.A.
 ZIP: 20006
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.5 in, 1.44MB
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS, Version 5.0
 SOFTWARE: ASCII
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/315,850
 FILING DATE:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/08/904,686
 FILING DATE: 01-AUG-1997
 APPLICATION NUMBER: US/08/324,977
 FILING DATE: 18-OCT-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 2-167466
 FILING DATE: 25-JUN-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 2-230921
 FILING DATE: 31-AUG-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 2-305605
 FILING DATE: 09-NOV-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/099,706
 FILING DATE: 30-JUL-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/769,996
 FILING DATE: 02-OCT-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/635,451
 FILING DATE: 28-DEC-1990
 ATTORNEY/AGENT INFORMATION:
 NAME: Mcleland, Le-Nhung
 REGISTRATION NUMBER: 31,541
 REFERENCE/DOCKET NUMBER: 9007036
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 659-2930
 TELEFAX: (202) 887-0357
 INFORMATION FOR SEQ ID NO: 31:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 7917 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA from genomic RNA
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..7862
 NAME/KEY: misc_feature
 LOCATION: 1..7916
 OTHER INFORMATION: /note= "sequence = 1500 - 9416 of
 OTHER INFORMATION: SEQ ID NO: 1"
 US-09-315-850-31

Query Match 87.9%; Score 1796.6; DB 3; Length 7917;
 Best Local Similarity 92.5%; Pred. No. 0;
 Matches 1889; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

QY 1 TGGGAGGGGCTTTTCAACAGGCTTCAACCGCATGATGCCACTTCTGTGCCAACAAG 60
 DB 3505 TGGGAGAGTGTCTTCAACAGGCTTCAACCGCATATATGATGCAACACTTCTGTGCCAACAAG 3564
 QY 61 CAGGAGAGAGCAACTTCCCTACTGATGAGGTACCAAGGCTAATGTGTGCGCTAGAGGC 120

DB 3565 CAGGAGAGAGCAACTTCCCTACTGATGAGGTACCAAGGCTAATGTGTGCGCTAGAGGC 3624
 QY 121 CAGGAGAGAGCAACTTCCCTACTGATGAGGTACCAAGGCTAATGTGTGCGCTAGAGGC 180
 DB 3625 CAGGAGAGAGCAACTTCCCTACTGATGAGGTACCAAGGCTAATGTGTGCGCTAGAGGC 3684
 QY 181 CTGCGGAGGAGCAACACCTTGTGATAGGCTGGAGCCGCTCCAAAGAGGTCAACCTC 240
 DB 3685 CTGCGAGGAGCAACACCTTGTGATAGGCTGGAGCCGCTCCAAAGAGGTCAACCTC 3744
 QY 241 ACACACCCCATTAACCAATTCATCATGAGCATGATGACCGACCTGAGAGTGTGACAG 300
 DB 3745 ACCACCCCATTAACCAATTCATCATGAGCATGATGACCGACCTGAGAGTGTGACAG 3804
 QY 301 AGCACTGGAGTGTGATGAGGCGGAGTCTTGAAGCTTGGCTGCGATTTGCTTGACAA 360
 DB 3805 AGCACTGGAGTGTGATGAGGCGGAGTCTTGAAGCTTGGCTGCGATTTGCTTGACAA 3864
 QY 361 GGAGAGGTGATGATGAGGCTGATGATCATCTTGTCCGGGCGGCGCTATTGTTCCGAG 420
 DB 3865 GGAGAGGTGATGATGAGGCTGATGATCATCTTGTCCGGGCGGCGCTATTGTTCCGAG 3924
 QY 421 AGGAGAGTCTCTACAGAGATTCATGATGATGAAAGTGGCGCTGACACCTCCCTAC 480
 DB 3925 AGGAGAGTCTCTACAGAGATTCATGATGATGAAAGTGGCGCTGACACCTCCCTAC 3984
 QY 481 ATGAGAGAGGATGAGAGTCTCCCGACGATTCAGCAAAAGAGCTCGGTTGCTGACAG 540
 DB 3985 ATGAGAGAGGATGAGAGTCTCCCGACGATTCAGCAAAAGAGCTCGGTTGCTGACAA 4044
 QY 541 ACAGCACAACAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 600
 DB 4045 ACAGCACAACAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4104
 QY 601 GAGACCTTCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 660
 DB 4105 GAGACCTTCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4164
 QY 661 TTGTCACTCTGCTGAGGAGTCCCGGATTCATCATGATGAGGAGGAGGAGGAGGAGGAG 720
 DB 4165 TTATCACTCTGCTGAGGAGTCCCGGATTCATCATGATGAGGAGGAGGAGGAGGAGGAG 4224
 QY 721 ACTAGCCGCTACACACCAATCTACCCGCTTCAATCTGAGGAGGAGGAGGAGGAGGAGG 780
 DB 4225 ACTAGCCGCTACACACCAATCTACCCGCTTCAATCTGAGGAGGAGGAGGAGGAGGAGG 4284
 QY 781 GCCCACTGCTCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 840
 DB 4285 GCCCACTGCTCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4344
 QY 841 GCTGTTGAGAGATAGGCTTGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 900
 DB 4345 GCTGTTGAGAGATAGGCTTGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4404
 QY 901 GAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 960
 DB 4405 GAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4464
 QY 961 GACCTGTTAACTTACTCCCTGCAATCTCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAG 1020
 DB 4465 GACCTGTTAACTTACTCCCTGCAATCTCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAG 4524
 QY 1081 CGGCTGATAGGCTTGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1140
 DB 4585 CGGCTGATAGGCTTGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4644
 QY 1141 AGCGAGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1200

```

Db 4645 AGCGAGCGCGGAGCGGCTGTCTACGATGCTCTCCAGCCTTACCATCTACGCTGCTG 4704
Qy 1201 AAGAGGCTCCACGAGTGAATTAACGAGGAGCTCTCCAGCGCTGCTCGGCTGAGGTA 1260
Db 4705 AAAAGGCTCCACGAGTGAATTAACGAGGAGCTCTCCAGCGCTGCTCGGCTGAGGTA 4764
Qy 1261 AGGAGTGTGGGAGTGAATTAACGAGGAGCTCTCCAGCGCTGCTCGGCTGAGGTA 1320
Db 4765 AGGAGTGTGGGAGTGAATTAACGAGGAGCTCTCCAGCGCTGCTCGGCTGAGGTA 4824
Qy 1321 AAGCTCCGCGCGGAGTGAATTAACGAGGAGCTCTCCAGCGCTGCTCGGCTGAGGTA 1380
Db 4825 AAGCTCCGCGCGGAGTGAATTAACGAGGAGCTCTCCAGCGCTGCTCGGCTGAGGTA 4884
Qy 1381 GTCTGGCGGGGAGAGCGGATCATGAGCAACCTCTGATGTGAGCAAGATCAACCGGA 1440
Db 4885 GTCTGGCGGGGAGAGCGGATCATGAGCAACCTCTGATGTGAGCAAGATCAACCGGA 4944
Qy 1441 CATGTCAAAAAGGTTCAATGAGATCGTTGGGCTTAAGACCTGTATGTAATCATGTGGCAT 1500
Db 4945 CATGTCAAAAAGGTTCAATGAGATCGTTGGGCTTAAGACCTGTATGTAATCATGTGGCAT 5004
Qy 1501 GGAACATTTCCCATCAAGCATACGACGAGGCGCTGACAGCGCTCTCCAGCGGCAAC 1560
Db 5005 GGAACATTTCCCATCAAGCATACGACGAGGCGCTGACAGCGCTCTCCAGCGGCAAC 5064
Qy 1561 TATTCAGAGGCGCTGTGGCGGGGTGCTGCTGAGAGTACGTTGAGGTTACGCGGGTGAGG 1620
Db 5065 TATTCAGAGGCGCTGTGGCGGGGTGCTGCTGAGAGTACGTTGAGGTTACGCGGGTGAGG 5124
Qy 1621 GATTTCCATACGTAACGAGCATGACCACTGACAAAGTAAATGCCCGTGCAGGTTCCA 1680
Db 5125 GATTTCCATACGTAACGAGCATGACCACTGACAAAGTAAATGCCCGTGCAGGTTCCA 5184
Qy 1681 GCCCGGAAATTTCTTCAAGAGTGAATGGGGGCGGCTGACAGGTAAGGCTCGGCGTGC 1740
Db 5185 GCCCGGAAATTTCTTCAAGAGTGAATGGGGGCGGCTGACAGGTAAGGCTCGGCGTGC 5244
Qy 1741 AAACCTCTCTACGAGGAGGAGTCAATTCAGTCCGAGCTCAACCAATACCTGTTGGG 1800
Db 5245 AGGCTCTCTCTACGAGGAGGAGTCAATTCAGTCCGAGCTCAACCAATACCTGTTGGG 5304
Qy 1801 TCGAGCTCCATGCGAGCGGCGGACCGGATGTAGAGTGTCTCATCTTCCATGCTCACGAC 1860
Db 5305 TCGAGCTCCATGCGAGCGGCGGACCGGATGTAGAGTGTCTCATCTTCCATGCTCACGAC 5364
Qy 1861 CCCTCCCATACAGAGGAGGAGGCTGAGCGGAGGCTGAGCGGAGGCTGAGCGGAGGCTG 1920
Db 5365 CCCTCCCATACAGAGGAGGAGGCTGAGCGGAGGCTGAGCGGAGGCTGAGCGGAGGCTG 5424
Qy 1921 TTGGCAGGCTCTTACGAGTGTGTGCGGCTTCTCGAAGGCGCATACATTAAC 1980
Db 5425 TTGGCAGGCTCTTACGAGTGTGTGCGGCTTCTCGAAGGCGCATACATTAAC 5484
Qy 1981 CAAAATGACTTCCAGAGCGCTGACCTCATGAGGCGCAACCTCTGTGCGGCGATGAGATG 2040
Db 5485 CACCATGTCTCTCCGAGCGCTGACCTCATGAGGCGCAACCTCTGTGCGGCGAGAGATG 5544
Qy 2041 GGC 2043
Db 5545 GGC 5547

```

```

RESUT: 11
US-08-324-977-13
; Sequence 13, Application US/08324977
; Patent No. 5747339

```

```

; GENERAL INFORMATION:
; APPLICANT: OKAYAMA, Hiroto
; APPLICANT: FUKE, Isao
; APPLICANT: MORI, Chisato
; APPLICANT: TAKAMIZAWA, Akahisa
; APPLICANT: YOSHIDA, Iwao

```

```

; TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC
; TITLE OF INVENTION: CDNA AND ANTIGEN POLYPEPTIDE
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Armstrong, Westerman, Hattori, McJeland &
; ADDRESS: Naughton
; STREET: 1725 K St. N.W. Suite 1000
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 in, 1.44MB
; OPERATING SYSTEM: PC-DOS/MS-DOS, Version 5.0
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/324,977
; FILING DATE: 18-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-167466
; FILING DATE: 25-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-230921
; FILING DATE: 31-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-305605
; FILING DATE: 09-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/099,706
; FILING DATE: 30-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/769,996
; FILING DATE: 02-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/635,451
; FILING DATE: 28-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Stevens-Smith, Theresa M.
; REGISTRATION NUMBER: 36,281
; REFERENCE/DOCKET NUMBER: 900703D
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 659-2930
; TELEFAX: (202) 887-0357
; TELEX: 440142
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9030 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA from genomic RNA
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1..9030
; OTHER INFORMATION: /note="sequence = 333 - 9362 of
; OTHER INFORMATION: SEQ ID NO: 1"
; NAME/KEY: CDS
; LOCATION: 1..9030
; US-08-324-977-13

```

```

Query Match 87.9%; Score 1796.6; DB 1; Length 9030;
Best Local Similarity 92.5%; Pred. No. 0;
Matches 1889; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

```

```

Qy 1 TTGGAGGCGGCTTCTACAGGCTTACCCAGTGAATGCCCACTTCTGTGCTCAACCAAG 60
Db 4672 TTGGAGGCGGCTTCTACAGGCTTACCCAGTGAATGCCCACTTCTGTGCTCAACCAAG 4731
Qy 61 CAGGAGAGGACAACTTCCCTCACTGAGGCGTACAGGCTACTGTGTGCGGCTAGGCGC 120
Db 4732 CAGGAGAGGACAACTTCCCTCACTGAGGCGTACCAAGGCGGCTGTGCGGCT 4791

```

QY 121 CAGGCCCACTTCATCATGAGATCAATGTGAAAGTGTCTCATACGCTAAAGCTACT 180
 Db 4792 CAGGCCCACTTCATCATGAGATCAATGTGAAAGTGTCTCATACGCTAAAGCTACT 4851
 QY 181 CTGCGCGGGCCCAACCCCTTGTGTATGAGCTGGAGCCGCTCCAAAACAGGTCACTC 240
 Db 4852 CTGCGCGGGCCCAACCCCTTGTGTATGAGCTGGAGCCGCTCCAAAACAGGTCACTC 4911
 QY 241 ACACGCCCATTAACCAATTCATCATGAGATCAATGTGAAAGTGTCTCATACGCTAAAGCTACT 300
 Db 4912 ACACGCCCATTAACCAATTCATCATGAGATCAATGTGAAAGTGTCTCATACGCTAAAGCTACT 4971
 QY 301 AGCACTGGGTGCTGTGGGCGGGGCTCTTGAGAGCTGTGAGCTGTGAGCTGTGAGCA 360
 Db 4972 AGCACTGGGTGCTGTGGGCGGGGCTCTTGAGAGCTGTGAGCTGTGAGCTGTGAGCA 5031
 QY 361 GGCAGCTGTGATTTGTGGTATGAGATCATTTGTCCGGGCGGCTATTTGTCCGAC 420
 Db 5032 GGCAGCTGTGATTTGTGGTATGAGATCATTTGTCCGGGCGGCTATTTGTCCGAC 5091
 QY 421 AGGGAAGTCTCTTACAGAGATTTGATGAGATGAGAGATGAGAGATGAGAGATGAGAGAT 480
 Db 5092 AGGGAAGTCTCTTACAGAGATTTGATGAGATGAGAGATGAGAGATGAGAGATGAGAGAT 5151
 QY 481 ATCGAGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 540
 Db 5152 ATCGAGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 5211
 QY 541 ACAGCCCAACCAAG 600
 Db 5212 ACAGCCCAACCAAG 5271
 QY 601 GAGACTCTTGTGGGCGGAAACATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 660
 Db 5272 GAGACTCTTGTGGGCGGAAACATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 5331
 QY 661 TTGTCACTCTGCTGGGAAATCCCGGATGATGATGATGATGATGATGATGATGATGATGAT 720
 Db 5332 TTGTCACTCTGCTGGGAAATCCCGGATGATGATGATGATGATGATGATGATGATGATGAT 5391
 QY 721 ACTAGCCGCTCACCAACCAATTCATCTCTGCTTAAATCATCTGGGAGATGAGATGAGAT 780
 Db 5392 ACTAGCCGCTCACCAACCAATTCATCTCTGCTTAAATCATCTGGGAGATGAGATGAGAT 5451
 QY 781 GCCCACTCTGCTGGGAAATCCCGGATGATGATGATGATGATGATGATGATGATGATGAT 840
 Db 5452 GCCCACTCTGCTGGGAAATCCCGGATGATGATGATGATGATGATGATGATGATGATGAT 5511
 QY 841 GCTGTGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 900
 Db 5512 GCTGTGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 5571
 QY 901 GAGAGTGGAG 960
 Db 5572 GAGAGTGGAG 5631
 QY 961 GACCTGTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1020
 Db 5632 GACCTGTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5691
 QY 1021 TGCGAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1080
 Db 5692 TGCGAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5751
 QY 1081 CGGCTGATAG 1140
 Db 5752 CGGCTGATAG 5811
 QY 1141 AGCGAG 1200
 Db 5812 AGCGAG 5871

QY 1201 AAGAGCTTCACAGATGATTAACAGAGATGCTCTCCAGCCCTGCTCCGGCTGTGGCTTA 1260
 Db 5872 AAGAGCTTCACAGATGATTAACAGAGATGCTCTCCAGCCCTGCTCCGGCTGTGGCTTA 5931
 QY 1261 AAGAGTGTGGAGATGATTAACAGAGATGCTCTCCAGCCCTGCTCCGGCTGTGGCTTA 1320
 Db 5932 AAGAGTGTGGAGATGATTAACAGAGATGCTCTCCAGCCCTGCTCCGGCTGTGGCTTA 5991
 QY 1321 AAGCTCTGCGGAGATTAACAGAGATGCTCTCCAGCCCTGCTCCGGCTGTGGCTTA 1380
 Db 5992 AAGCTCTGCGGAGATTAACAGAGATGCTCTCCAGCCCTGCTCCGGCTGTGGCTTA 6051
 QY 1381 GTCTGCGGAGAGAGAGAGATGATTAACAGAGATGCTCTCCAGCCCTGCTCCGGCTGTGGCTTA 1440
 Db 6052 GTCTGCGGAGAGAGAGAGATGATTAACAGAGATGCTCTCCAGCCCTGCTCCGGCTGTGGCTTA 6111
 QY 1441 CATGTCAAAAACGTTCCATGAGATGCTGTGGGCTTAAGACTGTATGATTAATGTGGCAT 1500
 Db 6112 CATGTCAAAAACGTTCCATGAGATGCTGTGGGCTTAAGACTGTATGATTAATGTGGCAT 6171
 QY 1501 GGAACATTCCTCATGAGATGATTAACAGAGATGCTCTCCAGCCCTGCTCCGGCTGTGGCTTA 1560
 Db 6172 GGAACATTCCTCATGAGATGATTAACAGAGATGCTCTCCAGCCCTGCTCCGGCTGTGGCTTA 6231
 QY 1561 TATTCAGGAGAGAGAGAGATGATTAACAGAGATGCTCTCCAGCCCTGCTCCGGCTGTGGCTTA 1620
 Db 6232 TATTCAGGAGAGAGAGAGATGATTAACAGAGATGCTCTCCAGCCCTGCTCCGGCTGTGGCTTA 6291
 QY 1621 GATTTCCATGATGATGATTAACAGAGATGCTCTCCAGCCCTGCTCCGGCTGTGGCTTA 1680
 Db 6292 GATTTCCATGATGATGATTAACAGAGATGCTCTCCAGCCCTGCTCCGGCTGTGGCTTA 6351
 QY 1681 GCGCCGGAATCTTCAAGAGATGATTAACAGAGATGCTCTCCAGCCCTGCTCCGGCTGTGGCTTA 1740
 Db 6352 GCGCCGGAATCTTCAAGAGATGATTAACAGAGATGCTCTCCAGCCCTGCTCCGGCTGTGGCTTA 6411
 QY 1741 AAACCTCTCTCAAGAGAGAGATGATTAACAGAGATGCTCTCCAGCCCTGCTCCGGCTGTGGCTTA 1800
 Db 6412 AAACCTCTCTCAAGAGAGAGATGATTAACAGAGATGCTCTCCAGCCCTGCTCCGGCTGTGGCTTA 6471
 QY 1801 TCGAGCTCCATGAGAGAGAGATGATTAACAGAGATGCTCTCCAGCCCTGCTCCGGCTGTGGCTTA 1860
 Db 6472 TCGAGCTCCATGAGAGAGAGATGATTAACAGAGATGCTCTCCAGCCCTGCTCCGGCTGTGGCTTA 6531
 QY 1861 CCCTCCCATGAGAGAGAGATGATTAACAGAGATGCTCTCCAGCCCTGCTCCGGCTGTGGCTTA 1920
 Db 6532 CCCTCCCATGAGAGAGAGATGATTAACAGAGATGCTCTCCAGCCCTGCTCCGGCTGTGGCTTA 6591
 QY 1921 TTGGCAGCTCTTCAAGAGAGAGATGATTAACAGAGATGCTCTCCAGCCCTGCTCCGGCTGTGGCTTA 1980
 Db 6592 TTGGCAGCTCTTCAAGAGAGAGATGATTAACAGAGATGCTCTCCAGCCCTGCTCCGGCTGTGGCTTA 6651
 QY 1981 CAAATGATTCACAGAGAGAGATGATTAACAGAGATGCTCTCCAGCCCTGCTCCGGCTGTGGCTTA 2040
 Db 6652 CAAATGATTCACAGAGAGAGATGATTAACAGAGATGCTCTCCAGCCCTGCTCCGGCTGTGGCTTA 6711
 QY 2041 GGC 2043
 Db 6712 GGC 6714

RESULT 12
 US-08-384-616-13
 ; Sequence 13, Application US/08384616
 ; Patent No. 5847101
 ; GENERAL INFORMATION:
 ; APPLICANT: OKAYAMA, Hiroto
 ; APPLICANT: FUKE, Iseo
 ; APPLICANT: MORI, Chisato
 ; APPLICANT: TAKAMIZAWA, Akahisa
 ; APPLICANT: YOSHIDA, Iwao
 ; TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC
 ; TITLE OF INVENTION: CDNA AND ANTIGEN POLYPEPTIDE

NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSSEE: Armstrong, Westerman, Hattori, Mclelland &
ADDRESSSEE: Naughton
STREET: 1725 K St. N.W. Suite 1000
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 in, 1.44Mb
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS, Version 5.0
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/384,616
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/769,996
FILING DATE: 02-OCT-1991
APPLICATION NUMBER: JP 2-167466
FILING DATE: 25-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 2-230921
FILING DATE: 31-AUG-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 2-305605
FILING DATE: 09-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/635,451
FILING DATE: 28-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: Stevens-Smith, Theresa M.
REGISTRATION NUMBER: 36,281
REFERENCE/DOCKET NUMBER: 900703B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 659-2930
TELEFAX: (202) 887-0357
TELEX: 440142
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 9030 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA from genomic RNA
FEATURE:
NAME/KEY: misc feature
LOCATION: 1..9030 /note: "sequence = 333 - 9362 of
OTHER INFORMATION: SEQ ID NO: 1"
FEATURE:
NAME/KEY: CDS
LOCATION: 1..9030
US-08-384-616-13

Query Match 87.9%; Score 1796.6; DB 2; Length 9030;
Best Local Similarity 92.5%; Pred. No. 0; Mismatches 154; Indels 0; Gaps 0;
Matches 1889; Conservative 0;

QY 1 TGGAGAGGCGCTTCAACAGGCTTCAACGAGGCTTCTGTCCTGCTCCAAACAAAG 60
DB 4672 TGGAGAGGCTTCAACAGGCTTCAACGAGGCTTCTGTCCTGCTCCAAACAAAG 4731
QY 61 CAGGAGAGGAGCACTTCCCTTCACTGCTGAGGCTTCAACGAGGCTTCTGTCCTGAGGCC 120
DB 4732 CAGGAGAGGAGCACTTCCCTTCACTGCTGAGGCTTCAACGAGGCTTCTGTCCTGAGGCC 4791
QY 121 CAGGAGGAGGCTTCACTGAGGCTTCAACGAGGCTTCTGTCCTGAGGCTTCAACGAGGCTTCACT 180
DB 4792 CAGGAGGAGGCTTCACTGAGGCTTCAACGAGGCTTCTGTCCTGAGGCTTCAACGAGGCTTCACT 4851

QY 181 CTGCGCGGGCCCAACACCTTGTCTATAGCTGGAGCCGCTCAAAACGAGTCAACCTC 240
DB 4852 CTGAGCGGGCCCAACACCTTGTCTATAGCTGGAGCCGCTCAAAACGAGTCAACCTC 4911
QY 241 ACACACCCATTAACCAATTCATCATGAGCATGATGTCAGCCGACCTGAGAGTGTACG 300
DB 4912 ACCACCCCATTAACCAATTCATCATGAGCATGATGTCAGCCGACCTGAGAGTGTACCT 4971
QY 301 AGCACCTGGGTGTGTGGGGCGGGGCTTGCACCTGTGCGTATTCCTTGACAA 360
DB 4972 AGCACCTGGGTGTGTGGGGCGGGGCTTGCACCTGTGCGGATTCCTTGACAA 5031
QY 361 GGCAGCGGTGTGTGTGGGGCGGGGCTTGCACCTGTGCGGATTCCTTGACAA 420
DB 5032 GGCAGGTGTGTGTGTGGGGCGGGGCTTGCACCTGTGCGGATTCCTTGACAA 5091
QY 421 AGGGAAGTCTCTACCAAGGATTCGATGAGATGGAAGAGTGGCGCTGCACTCCCTTAC 480
DB 5092 AGGGAAGTCTCTACCAAGGATTCGATGAGATGGAAGAGTGGCGCTGCACTCCCTTAC 5151
QY 481 ATCGAGCGGGAATGCAAGCTTCCGAGAGCTTCAAGCAAAAGCGCTGCGGTGTGACG 540
DB 5152 ATCGAGCGGGAATGCAAGCTTCCGAGAGCTTCAAGCAAAAGCGCTGCGGTGTGACG 5211
QY 541 ACAGCCCAACGAAGCGGAGCGCGCTGCTCCGCTGGTGGAGTCCAGTGGCGAGCCCTT 600
DB 5212 ACAGCCCAACGAAGCGGAGCGCGCTGCTCCGCTGGTGGAGTCCAGTGGCGAGCCCTT 5271
QY 601 GAGACCTTGTGGGGAACACATGTGGAATTCATACAGCGGGAATACAGTACTTACAGAGC 660
DB 5272 GAGACCTTGTGGGGAACACATGTGGAATTCATACAGCGGGAATACAGTACTTACAGAGC 5331
QY 661 TTGTCACTTGTGCTGGGAATCCCGGATTCATCATGATGCGGTTCACAGCTGTGTC 720
DB 5332 TTGTCACTTGTGCTGGGAATCCCGGATTCATCATGATGCGGTTCACAGCTGTGTC 5391
QY 721 ACTAGCCGCTGACCAACCAATTCATCCTCTGCTTAAATCCGCGGGGAGATGGGTAGCC 780
DB 5392 ACCAGCCGCTGACCAACCAATTCATCCTCTGCTTAAATCCGCGGGGAGATGGGTAGCC 5451
QY 781 GCCCAACTCGCTCCCGAGTGTGCTTCACTTCTGAGGCGCGGATTCGTCGTGTCG 840
DB 5452 GCCCAACTCGCTCCCGAGTGTGCTTCACTTCTGAGGCGCGGATTCGTCGTGTCG 5511
QY 841 GCTGTGGCAGCATAGGCTTGGGAAGTGTGTCATCTTGTGCGGCTATGAGCA 900
DB 5512 GCTGTGGCAGCATAGGCTTGGGAAGTGTGTCATCTTGTGCGGCTATGAGCA 5571
QY 901 GAGTGGCAGGCGGCTGTGAGCTTAAAGTCAAGAGCGGCAATCCCTCCACGAG 960
DB 5572 GAGTGGCAGGCGGCTGTGAGCTTAAAGTCAAGAGCGGCAATCCCTCCACGAG 5631
QY 961 GACCTGTTAACTTACTCCCTGCACTCTCTCTCTGTCGCTGCTGCTGCTGCTGCTG 1020
DB 5632 GACCTGTTAACTTACTCCCTGCACTCTCTCTCTGTCGCTGCTGCTGCTGCTGCTG 5691
QY 1021 TGGCAGGAGATCTGCGTGGGAGTGGGCTCAAGGAGGAGGAGGCTGTCAGTGTGTAAC 1080
DB 5692 TGGCAGGAGATCTGCGTGGGAGTGGGCTCAAGGAGGAGGAGGCTGTCAGTGTGTAAC 5751
QY 1081 CGGCTGATAGGCTTGTGCTGCGGGGTAAACATGTTTCCCAACGCACTATGTGCAAG 1140
DB 5752 CGGCTGATAGGCTTGTGCTGCGGGGTAAACATGTTTCCCAACGCACTATGTGCTGAG 5811
QY 1141 AGCGAGCGCGGAGCAGTGTCACTGAGTTCGCTCCGACTTACTATACCCAACTGTG 1200
DB 5812 AGCGAGCGCGGAGCAGTGTCACTGAGTTCGCTCCGACTTACTATACCCAACTGTG 5871
QY 1201 AAGAGCTTCAACAGTGTATTAACGAGAGTGTTCACAGCCCTGTGCTGCTGCTGCTG 1260
DB 5872 AAGAGCTTCAACAGTGTATTAACGAGAGTGTTCACAGCCCTGTGCTGCTGCTGCTG 5931
QY 1261 AGGATGTTTGGAGCTGATATGACAGAGTTTGGCTGACTTCAAGACCTGGCTCCAGTCC 1320

```

Db 5932 AGGATGTTGGAGCTGATATGCAACGGGTGACTGACTTCAGAACCTGGCTCCAGTCC 5991
Qy 1321 AAGCTCCGCGCGCAATACCGGGAGTCCCTTTTCTCATGCAACGTGGGTACAAGGGG 1380
Db 5992 AAGCTCTGCGCGAGCTACCTGAGTCCCTTTTCTGTCGCAACGGGGTACAAGGGG 6051
Qy 1381 GTCTGGCGGGAGACGGGATCATGAGACCACTGCTCATGTGAGCAAGATCAACCGA 1440
Db 6052 GTCTGGCGGGAGACGGGATCATGCAACCACTGCTCATGTGAGCAAGATCAACCGA 6111
Qy 1441 CATGTCAAAAAGGTTTCATGAGATCGTTGGGCTTAAGACTGTAGTAACTGTGGCAT 1500
Db 6112 CATGTCAAAAAGGTTTCATGAGATCGTTGGGCTTAAGACTGTAGTAACTGTGGCAT 6171
Qy 1501 GGAACATTTCCCATGAAGCAACACACGGGCGCTGACGGCCCTCCACGCGCAAC 1560
Db 6172 GGAACATTTCCCATGAAGCAACACACGGGCGCTGACGGCCCTCCACGCGCAAC 6231
Qy 1561 TATTCAGGGGCGCTGTGGCGGGTGGCTGTGAGAGTACGTGAGAGTTACGGGGTGGG 1620
Db 6232 TATTCAGGGGCGCTGTGGCGGGTGGCTGTGAGAGTACGTGAGAGTTACGGGGTGGG 6291
Qy 1621 GATTTCACTACGTACGAGCATGACCACTGCAACGTAAGTGCCTGCGAGGTTCCG 1680
Db 6292 GATTTCACTACGTACGAGCATGACCACTGCAACGTAAGTGCCTGCGAGGTTCCG 6351
Qy 1681 GCCCGCAATTTTTCACAGAAATGAGTGGGGTGGCTGACAGTACGCTCCGGCTGTC 1740
Db 6352 GCTCTGAATTTCTTCGGAGGTGAGCGAGTGGCGGTTGCAAGGTACGCTCCGGCTGTC 6411
Qy 1741 AAACCTCTCTACGGGAGAGGTCACTTCAGGTCCGGGCTCAACCAATACCTGGTGGG 1800
Db 6412 AGGCTCTCTCTACGGGAGAGGTCACTTCAGGTCCGGGCTCAACCAATACCTGGTGGG 6471
Qy 1801 TCGAGCTCCCATGCGAGCGCCGCAACCGATGTAGAGTGTCACTTCCATGCTCACGAC 1860
Db 6472 TCGAGCTCCCATGCGAGCGCCGCAACCGATGTAGAGTGTCACTTCCATGCTCACGAC 6531
Qy 1861 CCCTCCCATCATCACAGCAGAGCGGCTTAAGCGGCGAGGCGGCTCTCCCTCC 1920
Db 6532 CCCTCCCATCATCACAGCAGAGCGGCTTAAGCGGCGAGGCGGCTCTCCCTCC 6591
Qy 1921 TTGGCACTCTTCAAGTGTGCTGAGTGTGCTGCTCTCTGCAAGCGCATTAATACC 1980
Db 6592 TTGGCACTCTTCAAGTGTGCTGAGTGTGCTGCTCTCTGCAAGCGCATTAATACC 6651
Qy 1981 CAAATGACTTCCAGAGCTGACTCATTCAGAGCCAACTCTGTGGCGGATGAGATG 2040
Db 6652 CACCATGTCTCTCCGAGCGCTGACTCATTCAGAGCCAACTCTGTGGCGGATGAGATG 6711
Qy 2041 GGC 2043
Db 6712 GGC 6714

```

RESULT 13

US-08-904-686A-13
 ; Sequence 13, Application US/08904686A
 ; Patent No. 5998130

GENERAL INFORMATION:

APPLICANT: OKAYAMA, Hioto
 APPLICANT: FUKE, Iwao
 APPLICANT: MORI, Chisato
 APPLICANT: TAKAMIZAWA, Akahisa
 APPLICANT: YOSHIDA, Iwao
 TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC
 TITLE OF INVENTION: CDNA AND ANTIGEN POLYPEPTIDE
 NUMBER OF SEQUENCES: 50
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Armstrong, Westerman, Hattori, Mclelland &
 ADDRESSEE: Naughton
 STREET: 1725 K St. N.W. Suite 1000

```

CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Discrete, 3.5 in, 1.44Mb
OPERATING SYSTEM: PC-DOS/MS-DOS, Version 5.0
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/904,686A
FILING DATE: 01-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 2-230921
FILING DATE: 31-AUG-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 2-167466
FILING DATE: 25-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 2-305605
FILING DATE: 09-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/099,706
FILING DATE: 30-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/769,996
FILING DATE: 02-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/635,451
FILING DATE: 28-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: Mclelland, Le-Nung
REGISTRATION NUMBER: 31,541
REFERENCE/DOCKET NUMBER: 900703G
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 659-2930
TELEFAX: (202) 887-0357
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 9030 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA from genomic RNA
FEATURE:
NAME/KEY: misc feature
LOCATION: 1..9030
OTHER INFORMATION: /note: "sequence = 333 - 9362 of
FEATURE:
NAME/KEY: CDS
LOCATION: 1..9030
US-08-904-686A-13

```

Query Match 87.9%; Score 1796.6; DB 2; Length 9030;
 Best Local Similarity 92.5%; Pred. No. 0;
 Matches 1889; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

```

Qy 1 TGGGAGGCGCTTTACAGGCGCTCAACGATGATGCGCACTTCTGCAACCAAG 60
Db 4672 TGGGAGAGTGTTCACAGGCGCTCAACATATGATGACACTTCTTGTCCAGACCAAG 4731
Qy 61 CAGGAGAGAGCAACTTCCCTACCTGTGGCGTACAGAGCTACTGTGTGCGCTAGGGCC 120
Db 4732 CAGGAGAGAGCAACTTCCCTACCTGTGTGATATCAAGCAGGTGTGGCGCGGCT 4791
Qy 121 CAGGCCCACTTCATCATGAGATGAATGTGAAAGTGTTCATACGGCTAAAGCTTACT 180
Db 4792 CAGGCCCACTTCATCATGAGATGAATGTGAAAGTGTTCATACGGCTAAAGCTTACG 4851

```

QY 181 CTGCGGGGCCAACCCTTGCTGTATAGCTGGAGCCGCTCCAAACGAGTCAACCTC 240
 Db 4852 CTGACAGGGCCCAACCCCTTGCTGTACAGAGCTGGAGCGCTCCAAATGAGGTACCCCTC 4911
 QY 241 ACACACCCCATTAACCAATTATCATGTGGATGCAATGTCAAGCCGACTGAGGTCTCTACG 300
 Db 4912 ACCCAACCCCATTAACCAATTATCATGTGGATGCAATGTCTGACCTGAGAGGTCTCTCACT 4971
 QY 301 AGCAGCTGGGATGCTGGTGGGCGGGGCTCTGAGCTCTGAGCTGCTTAATTTGCTTGACAACA 360
 Db 4972 AGCAGCTGGGATGCTGGTGGGCGGGGCTCTGAGCTCTGAGCTGCTTAATTTGCTTGACAACA 5031
 QY 361 GGCAGCGTGTCTTACAGAGATTTGAGATCATTTGTCCGCGCGCGGCTATTGTTCCGAC 420
 Db 5032 GGCAGTGTGTCTTACAGAGATTTGAGATCATTTGTCCGCGCGCGGCTATTGTTCCGAC 5091
 QY 421 AGGGAAGTCTTACAGAGATTTGAGATCATTTGTCCGCGCGCGGCTATTGTTCCGAC 480
 Db 5092 AGGGAAGTCTTACAGAGATTTGAGATCATTTGTCCGCGCGCGGCTATTGTTCCGAC 5151
 QY 481 ATCGAGCAGGAATGAGCTCGCGAGAGTTCAAGCAAAAAGCGCTCGGATTGCTGACG 540
 Db 5152 ATCGAGCAGGAATGAGCTCGCGAGAGTTCAAGCAAAAAGCGCTCGGATTGCTGACG 5211
 QY 541 ACAGCCACCAAGCAAGCGAGGCGCTGCTCCGTTGTTGAGTCCAGTGGCGAGCCCTT 600
 Db 5212 ACAGCCACCAAGCAAGCGAGGCGCTGCTCCGTTGTTGAGTCCAGTGGCGAGCCCTT 5271
 QY 601 GAGACCTTCTGGGCGAAACACATGTGGAATCTTCAACGCGGATACAGTACTTACAGAGC 660
 Db 5272 GAGACCTTCTGGGCGAAACACATGTGGAATCTTCAACGCGGATACAGTACTTACAGAGC 5331
 QY 661 TTGTTCACCTCTGCTGGGAAATCCCGGATGTGATCACTGATGGCGTTCAACAGCTCTGTG 720
 Db 5332 TTATTCACCTCTGCTGGGAAATCCCGGATGTGATCACTGATGGCGTTCAACAGCTCTGTG 5391
 QY 721 ACTAGCCGCTCACACCCCAATCTACCTCTGCTTAACATCTGGGGGATGGGTAGCC 780
 Db 5392 ACTAGCCGCTCACACCCCAATCTACCTCTGCTTAACATCTGGGGGATGGGTAGCC 5451
 QY 781 GCCCACTCTGCTCCCGGAGTGTCTTCAAGTCTTCAAGGCGCGGCTTCTGTCTGTG 840
 Db 5452 GCCCACTCTGCTCCCGGAGTGTCTTCAAGTCTTCAAGGCGCGGCTTCTGTCTGTG 5511
 QY 841 GCTGTGGAGCATAGGAGCTTGGAAAGGCTTGTGAGACATCTTGGCGGCTATGAGAGA 900
 Db 5512 GCTGTGGAGCATAGGAGCTTGGAAAGGCTTGTGAGACATCTTGGCGGCTATGAGAGA 5571
 QY 901 GGAAGTGGAGGCGCGCTGTGGCTTTAAGTCAATGAGCGGCGAAATGCCCTCCACGAG 960
 Db 5572 GGAAGTGGAGGCGCGCTGTGGCTTTAAGTCAATGAGCGGCGAAATGCCCTCCACGAG 5631
 QY 961 GACCTGTGTTAATTACTCTCTGCTCATCTCTCTGCTGCTGTGCTGGGGTCTGT 1020
 Db 5632 GACCTGTGTTAATTACTCTCTGCTCATCTCTCTGCTGCTGTGCTGGGGTCTGT 5691
 QY 1021 TGGCAGAGCATACTGCTGTGGAGTGGGTCCAGGGAGGGGGCTGTGAGTGAATGAAC 1080
 Db 5692 TGGCAGAGCATACTGCTGTGGAGTGGGTCCAGGGAGGGGGCTGTGAGTGAATGAAC 5751
 QY 1081 CGGCTGATAGCTTGGCTCGCGGGGTAACATGTTTCCCGCAGCAGCATATGTGCAGAG 1140
 Db 5752 CGGCTGATAGCTTGGCTCGCGGGGTAATGTTTCCCGCAGCAGCATATGTGCAGAG 5811
 QY 1141 AGCGAGCGCGCAGCAGTGTCACTCAATGATCTCTCTGAGCTTACTATCAACCACTGTTG 1200
 Db 5812 AGCGAGCGCGCAGCAGTGTCTCAAGATCTCTCTGAGCTTACTATCACTCACTGCTGCTG 5871
 QY 1201 AAGAGGCTCCACAGAGATTAAGAGAGCTGCTCCAGCGCGCTCGGCTCGGCTGAGGCTA 1260
 Db 5872 AAGAGGCTCCACAGAGATTAAGAGAGCTGCTCCAGCGCGCTCGGCTCGGCTGAGGCTA 5931
 QY 1261 AGGGAATGTTGGAGCTGATATGACAAGTGTGCTGACTTCAAGACCTGGCTCCAGTCC 1320

Db 5932 AGGGAATGTTGGAGCTGGAATATGACAGGTGTGACTGACTTCAAGACTGTGCTCCAGTCC 5991
 QY 1321 AAGCTCTCGCCGATTAACCGGAGTCCCTTTTCTCATGTCCAAAGTGGTTACAAGGGG 1380
 Db 5992 AAGCTCTCGCCGATTAACCGGAGTCCCTTTTCTCATGTCCAAAGTGGTTACAAGGGG 6051
 QY 1381 GTCTGGCGGGAGAGCGGATCATGACAGACACTGCTCATGTGAGAGCAGATCACCGGA 1440
 Db 6052 GTCTGGCGGGAGAGCGGATCATGACAGACACTGCTCATGTGAGAGCAGATCACCGGA 6111
 QY 1441 CATGTCAAAAACGGTTCCATGAGATGCTTGGGCTTAAGACTGTAGTAATATGTTGGAT 1500
 Db 6112 CATGTCAAAAACGGTTCCATGAGATGCTTGGGCTTAAGACTGTAGTAATATGTTGGAT 6171
 QY 1501 GGAACATTCCTCCATCAACGATACACCAAGCGGCGCCCTGACAGCCCTCCAGCGCAAC 1560
 Db 6172 GGAACATTCCTCCATCAACGATACACCAAGCGGCGCCCTGACAGCCCTCCAGCGCAAC 6231
 QY 1561 TATTCAGGGGCTGTGGCGGGTGGCTGCTGAGAGATACGTGAGATTACCGGGTGGG 1620
 Db 6232 TATTCAGGGGCTGTGGCGGGTGGCTGCTGAGAGATACGTGAGATTACCGGGTGGG 6291
 QY 1621 GATTTCACATCTGTCGACGATGACCACTGACCAACGTAATGCCCCGTGCTCAGGTTCCA 1680
 Db 6292 GATTTCACATCTGTCGACGATGACCACTGACCAACGTAATGCCCCGTGCTCAGGTTCCA 6351
 QY 1681 GCCCGCGAATCTTCAAGAGTGGATGGGGTGGCGGCTGACAGAGTACGCTCCGGCGTGC 1740
 Db 6352 GCTCTGAATCTTCTGCGAGAGTGGACGAGATGGCGGTTGACAGAGTACGCTCCGGCGTGC 6411
 QY 1741 AAACCTCTCTTACGGAGAGAGTCAATTCAGAGTGGGCTCAACCAATTAACCTGTTGGG 1800
 Db 6412 AAGCCTCTCTTACGGAGAGAGTCAATTCAGAGTGGGCTCAACCAATTAACCTGTTGGG 6471
 QY 1801 TCGAGCTCCCATGCGAGCCGGAACCGGATGTAGAGTGTCTCACTTCCATGCTACCGAC 1860
 Db 6472 TCGAGCTCCCATGCGAGCCGGAACCGGATGTAGAGTGTCTCACTTCCATGCTACCGAC 6531
 QY 1861 CCTTCCCATCAACAGCAGAGCGGCTAAGCGGCTGAGCGAGGGGCTTCCCGCTCC 1920
 Db 6532 CCTTCCCATCAACAGCAGAGCGGCTAAGCGGCTGAGCGAGGGGCTTCCCGCTCC 6591
 QY 1921 TTGGCAGCTCTTACAGTACGAGTGTGTCTGCGCTTCTCTGAAAGCGACATTAAC 1980
 Db 6592 TTGGCAGCTCTTACAGTACGAGTGTGTCTGCGCTTCTCTGAAAGCGACATTAAC 6651
 QY 1981 CAAATGACTTCCAGACGCTGACCTCATGAGGCGCAACTCTGTGCGGCTATGAGATG 2040
 Db 6652 CAAATGACTTCCAGACGCTGACCTCATGAGGCGCAACTCTGTGCGGCTATGAGATG 6711
 QY 2041 GGC 2043
 Db 6712 GGC 6714

RESULT 14
 us-09-315-850-13
 ; Sequence 13, Application US/09315850
 ; Patent No. 6217872
 ; GENERAL INFORMATION:
 ; APPLICANT: OKAYAMA, Hiroto
 ; APPLICANT: FUKE, Isao
 ; APPLICANT: MORI, Chisato
 ; APPLICANT: TAKAMIZAWA, Akahisa
 ; APPLICANT: YOSHIDA, Iwao
 ; TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC
 ; NUMBER OF SEQUENCES: 50
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSER: Armstrong, Westerman, Hattori, Mclelland &
 ; ADDRESSER: Naughton
 ; STREET: 1725 K St., N.W., Suite 1000


```

Db      5244 ACCACCCCATAAACAAATACATCATGATGATGATCGCTGACCTGAGAGGCTGCTACCT 5303
QY      301 AGCACCCTGGTGTCTGCTGGGCGGGGCTCTTGACGCTCTGCTGCGGATTTCTGGAACAA 360
Db      5304 AGCACCCTGGTGTCTGCTGGGCGGGGCTCTTGACGCTCTGCTGCGGATTTCTGGAACAA 5363
QY      361 GGCACGCTGTGATTTGTGGGTAGATCATCTTGTCCGGGCGGCGCTATTTGTTCCCGAC 420
Db      5364 GGCACGCTGTGATTTGTGGGTAGATCATCTTGTCCGGGAGCGGCGCATTTGTTCCCGAC 5423
QY      421 AGGGAAGTCTCTACCGAGAGTTGATGATGATGAAAGTGTGCGATCTGCGACCTCTTAC 480
Db      5424 AGGGAAGTCTCTCTACCGAGAGTTGATGATGATGAAAGTGTGCGATCTGCGACCTCTTAC 5483
QY      481 ATTCGACGAGGAATGAGCTCGCGGAGCACTTCAACCAAAAAGCGCTGGGTGCTGCGAG 540
Db      5484 ATTCGACGAGGAATGAGCTCGCGGAGCAATTCMAAGCAAAAAGCGCTGGGTGCTGCGAA 5543
QY      541 ACAGCCACCAAGCAAGCGAGCGGCTGCTCCGCTGGTGAAGTCCAAAGTGGCGAGCCCTT 600
Db      5544 ACAGCCACCAAGCAAGCGAGCGGCTGCTCCGCTGGTGAAGTCCAAAGTGGCGAGCCCTT 5603
QY      601 GAGACCTTCTGGGCGGAAACATATGTGAACTTTCATAGCGGGAATACAGTACTTACAGAGC 660
Db      5604 GAGACATTTCTGGGCGGAAACATATGTGAACTTTCATAGCGGGAATACAGTACTTACAGAGC 5663
QY      661 TTGTCCACTCTGCTGGGGAATCCCGGATTTGATCTCATGATGAGCGCTTCAAGCTCTGTC 720
Db      5664 TTATTCACCTCTGCTGGGGAATCCCGGATTTGATCTCATGATGAGCGCTTCAAGCTCTGTC 5723
QY      721 ACTAGCCGCTCAACCAACATCTACCTCTGCTTAAATCTCTGGGGGATGGTATGCG 780
Db      5724 ACTAGCCGCTCAACCAACATCTACCTCTGCTTAAATCTCTGGGGGATGGTATGCG 5783
QY      781 GCCCACTGCTCCCGGAGTGTGCTTCACTTTGATGAGCGCGCGCATTTGCTGTGCG 840
Db      5784 GCCCACTGCTCCCGGAGTGTGCTTCACTTTGATGAGCGCGCGCATTTGCTGTGCG 5843
QY      841 GCTGTGGGAGCATAGGCTTTGGGAAAGTGTGTTGATGATCTTGGCGGGCTATGAGACA 900
Db      5844 GCTGTGGGAGCATAGGCTTTGGGAAAGTGTGTTGATGATCTTGGCGGGCTATGAGACA 5903
QY      901 GAGAGTGGAGGCGGCTGTGAGCTTTAAGTATAGAGGCGGAAATGCGCTTCAACGAG 960
Db      5904 GAGAGTGGAGGCGGCTGTGAGCTTTAAGTATAGAGGCGGAAATGCGCTTCAACGAG 5963
QY      961 GACCTGTTAATCTACCTCCCTGCAATCTCTCTCTGATGAGCTGTGCTGTGCTGTG 1020
Db      5964 GACCTGTTAATCTACCTCCCTGCAATCTCTCTCTGATGAGCTGTGCTGTGCTGTG 6023
QY      1021 TGCAGAGCATATCTGCTGTGCGGAGCTGTGCGGAGAGGCGGCTGTGAGTGAATGAA 1080
Db      6024 TGTGACGCAATATCTGCTGTGCGGAGCTGTGCGGAGAGGCGGCTGTGAGTGAATGAA 6083
QY      1081 GGGCTGATAGGCTGTGCTGTGCGGAGGTAACATGTTTCCCGCAGCATATGAGCGAG 1140
Db      6084 GGGCTGATAGGCTGTGCTGTGCGGAGGTAACATGTTTCCCGCAGCATATGAGCGAG 6143
QY      1141 AGCGAGCGCGGAGCATGATCTCACTGATCTCTCCGACTTATCTATCAACCACTGTTG 1200
Db      6144 AGCGAGCGCGGAGCATGATCTCACTGATCTCTCCGACTTATCTATCACTGATGCTG 6203
QY      1201 AAGAGGCTCAACCATGATTAACAGAGACTGCTCAACGCGCTGCTCGGCTGTGGCTA 1260
Db      6204 AAGAGGCTCAACCATGATTAACAGAGACTGCTCAACGCGCTGCTCGGCTGTGGCTA 6263
QY      1261 AAGGATGTTTGGAGCTGATATGATGATGATGATGATGATGATGATGATGATGATGAT 1320
Db      6264 AAGGATGTTTGGAGCTGATATGATGATGATGATGATGATGATGATGATGATGATGAT 6323
QY      1321 AAGCTCTGCGGAGTATACCGGAGTCCCTTTTCTCATGCGCAAGTGGGTACAGGGG 1380

```

```

Db      6324 AAGCTCTGCGGAGTATACCGGAGTCCCTTTTCTGATGCCAAACGCGGTACAAAGGGA 6383
QY      1381 GTCTGGCGGGAGACCGCATCATGACAGACACTGCTCATGTGAGACAGATCACCGGA 1440
Db      6384 GTCTGGCGGGAGACCGCATCATGACAGACACTGCTCATGTGAGACAGATCACCGGA 6443
QY      1441 CATGTCMAAAACGGTTCATGAGATGCTTGGGCTTAAAGCTGTATGATATGTCAT 1500
Db      6444 CATGTCMAAAACGGTTCATGAGATGCTTGGGCTTAAAGCTGTATGATATGTCAT 6503
QY      1501 GGAACATTTCCCATCAACCATATACACAGAGGCGCTTGAACGCGCTTCCCGAGCCAAAC 1560
Db      6504 GGAACATTTCCCATCAACCATATACACAGAGGCGCTTGAACGCGCTTCCCGAGCCAAAC 6563
QY      1561 TATTCAGGAGCGTGTGCGGGTGTGCTGTGAGAGTATGATGAGAGTTACGCGGGTGGG 1620
Db      6564 TATTCAGGAGCGTGTGCGGGTGTGCTGTGAGAGTATGATGAGAGTTACGCGGGTGGG 6623
QY      1621 GATTTCCACTAGTACAGATGACCACTGACAAAGTAAATGCCCCGTGCGAGTTTCA 1680
Db      6624 GATTTCCACTAGTACAGATGACCACTGACAAAGTAAATGCCCCGTGCGAGTTTCCG 6683
QY      1681 GCCCGGAATTTCTTACAGAAAGTGTGAGGCTGCGCTGACAGATGAGCTCCGGGCTGC 1740
Db      6684 GCTTCTGAATTTCTTCTGAGAGGTGACAGAGAGTGTGACAGATGAGTACCTCCGGGCTGC 6743
QY      1741 AAACCTCTCTACGAGAGAGGTGACATTCAGAGTGGGCTCAACCAATATCTGTTGGG 1800
Db      6744 AGGCTTCTCTACGAGAGAGGTGACATTCAGAGTGGGCTCAACCAATATCTGTTGGG 6803
QY      1801 TGCAGCTCCATGCGAGGCGGAAACCGAGTGTAGAGTGTCACTTCCATGCTCACCGAG 1860
Db      6804 TGCAGCTCCATGCGAGGCGGAAACCGAGTGTAGAGTGTCACTTCCATGCTCACCGAG 6863
QY      1861 CCTTCCCATATACAGACAGAGACGGCTTAAGCGAGGCTGGCCAGGGGGTCTCCCTCTCC 1920
Db      6864 CCTTCCCATATACAGACAGAGACGGCTTAAGCGAGGCTGGCCAGGGGGTCTCCCTCTCC 6923
QY      1921 TTGGCCAGCTTTCAGCTAGCTAGTGTGTGCGCTTCTCGAAGGCGCATATATATAC 1980
Db      6924 TTGGCCAGCTTTCAGCTAGCTAGTGTGTGCGCTTCTCGAAGGCGCATATATATAC 6983
QY      1981 CAAATGACTTCCAGAGCGCTGACCTCATGAGAGCCAACTCTGTGCGGAGTGAATG 2040
Db      6984 CAAATGACTTCCAGAGCGCTGACCTCATGAGAGCCAACTCTGTGCGGAGTGAATG 7043
QY      2041 GGC 2043
Db      7044 GGC 7046

```

Search completed: February 25, 2005, 07:41:27
 Job time : 371 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2005, 03:07:49 / Search time 6005 Seconds
(without alignments)
12950.092 Million cell updates/sec

Title: US-09-664-363-20
2043
Sequence: 1 TGGGAGGCGCTCTTCACAGC.....TGTGGCGCATGATGAGCGC 2043

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues
Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
EST:
1: gb_ests1:
2: gb_ests2:
3: gb_hic:
4: gb_esc3:
5: gb_esc4:
6: gb_esc5:
7: gb_esc6:
8: gb_gsa1:
9: gb_gsa2:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|------------|---------------------|
| 1 | 49.2 | 2.4 | 925 | CNS0091P | AL053013 Drosophila |
| 2 | 48 | 2.3 | 552 | CF824031 | CF824031 EST701413 |
| 3 | 48 | 2.3 | 856 | CF824624 | CF824624 EST702006 |
| 4 | 47.6 | 2.3 | 938 | CF824954 | CF824954 EST702336 |
| 5 | 47 | 2.3 | 635 | BX905657 | BX905657 Leishman |
| 6 | 47 | 2.3 | 635 | LBAP018C06 | LBAP018C06 Leishman |
| 7 | 47 | 2.3 | 855 | CO015154 | CO015154 EST78536 |
| 8 | 47 | 2.3 | 909 | CF823313 | CF823313 EST700695 |
| 9 | 47 | 2.3 | 914 | CF826948 | CF826948 EST704330 |
| 10 | 47 | 2.3 | 923 | CF823378 | CF823378 EST700760 |
| 11 | 47 | 2.3 | 925 | CF825990 | CF825990 EST703372 |
| 12 | 46.6 | 2.3 | 925 | CNS0091P | AL053013 Drosophila |
| 13 | 46.2 | 2.3 | 645 | CNS01213 | AL01589 Drosophila |
| 14 | 45.8 | 2.2 | 881 | CF821532 | CF821532 EST698914 |
| 15 | 45.4 | 2.2 | 753 | CG855542 | CG855542 ZMMBC025 |
| 16 | 45.4 | 2.2 | 1101 | CNS01720 | AL107514 Drosophila |
| 17 | 45.2 | 2.2 | 1087 | BZ573831 | BZ573831 mah2_3381 |
| 18 | 45.2 | 2.2 | 1101 | CNS0178Y | AL108460 Drosophila |
| 19 | 43.2 | 2.1 | 966 | CNS0283T | AL210098 Tetradon |
| 20 | 43.2 | 2.1 | 971 | CNS04VB9 | AL308862 Tetradon |
| 21 | 43.2 | 2.1 | 1019 | CNS0411T | AL292286 Tetradon |
| 22 | 42.8 | 2.1 | 758 | CO015153 | CO015153 EST785535 |
| 23 | 42.6 | 2.1 | 786 | CF826564 | CF826564 EST703946 |
| 24 | 42.6 | 2.1 | 792 | CF821407 | CF821407 EST698789 |

| | | | | | | |
|------|------|-----|-----|---|----------|---------------------|
| C 25 | 42.6 | 2.1 | 794 | 7 | CF825686 | CF825686 EST703068 |
| C 26 | 42.6 | 2.1 | 799 | 7 | CF820066 | CF820066 EST697448 |
| C 27 | 42.6 | 2.1 | 800 | 7 | CO023701 | CO023701 EST788834 |
| C 28 | 42.6 | 2.1 | 808 | 7 | CF821886 | CF821886 EST699268 |
| C 29 | 42.6 | 2.1 | 822 | 7 | CF821953 | CF821953 EST699335 |
| C 30 | 42.6 | 2.1 | 830 | 7 | CO024108 | CO024108 EST789241 |
| C 31 | 42.6 | 2.1 | 834 | 7 | CF825132 | CF825132 EST702514 |
| C 32 | 42.6 | 2.1 | 841 | 7 | CF820947 | CF820947 EST698329 |
| C 33 | 42.6 | 2.1 | 841 | 7 | CF825355 | CF825355 EST702737 |
| C 34 | 42.6 | 2.1 | 849 | 7 | CF824541 | CF824541 EST701923 |
| C 35 | 42.6 | 2.1 | 851 | 7 | CO015200 | CO015200 EST785582 |
| C 36 | 42.6 | 2.1 | 854 | 7 | CF821753 | CF821753 EST699135 |
| C 37 | 42.6 | 2.1 | 857 | 7 | CF827683 | CF827683 EST705065 |
| C 38 | 42.6 | 2.1 | 864 | 7 | CF823454 | CF823454 EST700836 |
| C 39 | 42.6 | 2.1 | 868 | 7 | CF819880 | CF819880 EST697262 |
| C 40 | 42.6 | 2.1 | 868 | 7 | CF824362 | CF824362 EST701744 |
| C 41 | 42.6 | 2.1 | 876 | 7 | CF824955 | CF824955 EST702337 |
| C 42 | 42.4 | 2.1 | 935 | 9 | CNS006XK | AL066051 Drosophila |
| C 43 | 42.4 | 2.1 | 970 | 8 | BZ573821 | BZ573821 mah2_3377 |
| C 44 | 41.8 | 2.0 | 635 | 7 | CK975033 | CK975033 4106144 B |
| C 45 | 41.8 | 2.0 | 662 | 7 | CR454050 | CR454050 CR454050 |

ALIGNMENTS

RESULT 1
CNS0091P
LOCUS: CNS0091P 925 bp DNA linear GSS 03-JUN-1999
DEFINITION: Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR19P16 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION: AL053013.1 GI:4934461
VERSION: GSS.
KEYWORDS: Drosophila melanogaster (fruit fly)
SOURCE: Drosophila melanogaster
ORGANISM: Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE: Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage: BP 191 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr)
AUTHORS: - Web: www.genoscope.cns.fr
TITLE: Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP).
JOURNAL: The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammose in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

COMMENT

FEATURES
source
1..925
location/Qualifiers
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR19P16"
/clone_lib="RPCI-98"
/note="end : TET3"

ORIGIN

Query Match 2.4%; Score 49.2; DB 9; Length 925;
Best Local Similarity 13.9%; Pred. No. 0.17;

| | Matches | 51; Conservative | 161; Mismatches | 154; Indels | 0; Gaps | 0; |
|----|---------|---|-----------------|-------------|---------|----|
| Qy | 761 | TCTCTGGGGGAGATGGGTACCCGCCCAACTGCTCTCCCCCGACAGTCTCTTACAGTTTTCGAG | 820 | | | |
| Db | 553 | TTSSGSGYGKCCSSGSGSCSCSSCSGCCBCCGCCSSYCCSSBSSSKCS | 612 | | | |
| Qy | 821 | GCGCCGCGATTGCTGTCGTGCGCTGTGTGCACAGATAGGCGCTTGGAGAGGTGCTTGTGCACA | 880 | | | |
| Db | 613 | TSBSGSCCCSSKSVCTGCTSCSSSSGCSSSSSSTSSSTSSSTSSKSSSGSSSSSYTTS | 672 | | | |
| Qy | 881 | TCTTGGCGGGCTATGACAGACAGAGTGCACAGCGCGCTGTGCTTTAAGTCATAGACG | 940 | | | |
| Db | 673 | KSTSGSGSGSWACGGSGSTGCTSSSSSSSTSTSSSVSGSKSTBSGGBSSGSSS | 732 | | | |
| Qy | 941 | GCGAATCCCTCCACCGACGACCTGTGTTACTTCTCCCGACCATCTCTCTCTCTGAG | 1000 | | | |
| Db | 733 | SSSTSSBSCTSTSSSSSSSYSSSTCSCTCCCSYSSSTSSSTSSWTGSGSSBSV | 792 | | | |
| Qy | 1001 | CCCTGAGTGTGCGGGGTCTGTGTCGACGAGCATCTACGTGCGACATGTGGTCCAGCGAGAG | 1060 | | | |
| Db | 793 | GTSSSSDSTSTCCGCCCTMCCTCCSYMBCHYSTGCGSSSSSGKGGVTKCGCGCGSSS | 852 | | | |
| Qy | 1061 | GAGCTGTGCAGTGAATGAAACCGGCTGATAGCGTTGCGCTGCGGGGTAAACAATGTTTCC | 1120 | | | |
| Db | 853 | TNGMBGTSSACGSSSSSSSCSSSVSSSKSASASSSVSSSGSGVSSNESAASKSSSGSV | 912 | | | |
| Qy | 1121 | CCAGCG 1126 | | | | |
| Db | 913 | SSGSGS 918 | | | | |

RESULT 2
 CF824031/c
 LOCUS
 DEFINITION
 552 bp mRNA linear EST 01-APR-2004
 EST701413 Coccidioides posadasii saprobic phase cDNA library, 2 to
 4 kb Coccidioides posadasii cDNA clone CIDAP05 5' end, mRNA
 sequence.
 CF824031
 CF824031.1 GI:4593008
 EST.
 Coccidioides posadasii
 Coccidioides posadasii
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 Onygenales; mitosporic Onygenales; Coccidioides.
 1 (bases 1 to 552)
 Gardner, M.J. and Cole, G.T.
 Analysis of gene expression in Coccidioides posadasii mycelia and
 spherules via expressed sequence tags
 Unpublished (2003)
 Other_ESTs: EST701412
 Contact: Gardner MJ
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301 838 3519
 Fax: 301 838 0208
 Email: gardner@tigr.org
 Seq primer: MJ3 Reverse.
 location/Qualifiers
 1..552

ORIGIN

| | Query Match | 2.3%; | Score 48; | DB 7; | Length 552; |
|----|-----------------------|--|-----------------|-----------|-------------|
| | Best Local Similarity | 50.4%; | Pred. No. 0.33; | | |
| | Matches 117; | Conservative 0; | Mismatches 115; | Indels 0; | Gaps 0; |
| Qy | 797 | CCAGTCTCTCTTACGCTTTGCTAGAGCCCGCGCATTTGCTGTGCGGCTTTGGCAGCAATAG | 856 | | |
| Db | 522 | CCATGGCCGCGCGTCTTGTGTACCCGCGCTTGTGGCCATGTGTGTGTGTGGCCGCTGTGT | 463 | | |
| Qy | 857 | GCTTTGGGAAGAGTCTCTGTGTGACATCTTTGGCGGCGCTATGAGACAGATGTGCAGCGCGC | 916 | | |
| Db | 462 | CCATGCGCCGCGTCTTGTGTGCGCATATGTTCATATGACCGCGCGTGTGTGGCCGTGTGGC | 403 | | |
| Qy | 917 | TCTGTGSCCTTTAAGTTCATGAGCGGCGAATATCCCTCCACGAGAGACTGTGTTAACTTAC | 976 | | |
| Db | 402 | TTGTGTGCGCTGTGTTCATATGCGCGCGCTGTGTGGCCATATGTCCATATGCGCGCGCTGC | 343 | | |
| Qy | 977 | TCCCTGGCATCTCTCTCTGTATGCCCTGTGCTGTGAGGAGCTGTGTGCGCAGC | 1028 | | |
| Db | 342 | TTGTGTGCGCATATGTTCATATGCGCGCGCTGTGTGTGGCCGTGTGTGTTCATATGCC | 291 | | |

RESULT 3
 CF824624/c
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 FEATURES
 SOURCE

CF824624 856 bp mRNA linear EST 01-APR-2004
 EF8702006 Coccidioides posadasii saprobic phase cDNA library, 2 to
 4 kb Coccidioides posadasii cDNA clone CIDAS39 5' end, mRNA
 sequence.
 CF824624
 CF824624.1 GI:45930681
 EST.
 Coccidioides posadasii
 Coccidioides posadasii
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 Onygenales; mitosporic Onygenales; Coccidioides.
 1 (bases 1 to 856)
 Gardner, M.J. and Cole, G.T.
 Analysis of gene expression in Coccidioides posadasii mycelia and
 spherules via expressed sequence tags
 unpublished (2003)
 Contact: Gardner MJ
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301 838 3519
 Fax: 301 838 0208
 Email: Gardner@cigr.org
 Seq primer: M13 Reverse.
 Location/Qualifiers
 1..856

```

/db xref="taxon:199306"
/clone="CIDA39"
/dev stage="saprohic phase (mycelia)"
/lab_host="E. coli DH103, T1 phage resistant"
/clone_1ib="Coccidioides posadasii saprobic phase cDNA
library, 2 to 4 kb"
/note="Vector: pExpress 1; Site 1: Not I; Site 2: Eco RV;
Coccidioides posadasii saprobic phase cDNA library, size
fractionated cDNA 2 to 4 kb"

```

| | | | | |
|---------------------------|-------|----------------|----------|------------|
| Query Match | 2.3% | Score 48 | DB 7 | Length 856 |
| Best Local Similarity | 50.4% | Pred. No. 0.35 | | |
| Matches 117, Conservative | 0 | Mismatches 115 | Indels 0 | Gaps 0 |

| | | | |
|----|-----|---|-----|
| Qy | 797 | COAGNGCTTCAGCTTCGAGCGCGGCGATGCTGAGTGCAGCATAG | 856 |
| | | | |
| Db | 791 | CCATGCGCGCGTACTTGTGACCGCGCTTGTTGGCCATCTGCTTGCGCGTCGT | 732 |
| | | | |
| Qy | 857 | GCCTTGGAGAGTGGCTTGTGACATCTTGGCGGCGCTATGAGCAGAGCGAGCGCGC | 916 |
| | | | |
| Db | 731 | CCATGCGCGCGCGTGTGTGGCCATCATGTCCATGAGCGCGCGCTGTGTGGCGCGTCGTC | 672 |
| | | | |

QY 917 TCGTGACCTTTAAAGTCAATGAGCGGAAATGCCCTCCACGAGACCTGTAACTTAC 976
 DB 671 TTGTGCGCGTGTCTCCATGCGCCGCGCTGTGTGACATCATGTCATGCGCCGCTGC 612
 QY 977 TCCCTGCATCTCTCTCTGTGTCCTGTGTCGTGCGGGTGTGTGCGCAGC 1028
 DB 611 TTGTGCGCATCATGTCATGCGCGCGCTGTGTGCGCGCTGTGTGTCATGCGC 560
 RESULT 4
 CF824954 978 bp mRNA linear EST 01-APR-2004
 LOCUS EST702336 Coccidioides posadasii saprobic phase cDNA library, 2 to
 DEFINITION 4 kb Coccidioides posadasii cDNA clone CIDAU34 3' end, mRNA
 sequence.
 ACCESSION CF824954 GI:45931011
 VERSION CF824954.1
 KEYWORDS Coccidioides posadasii
 SOURCE Coccidioides posadasii
 ORGANISM Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 Onygenales; Mitosporic Onygenales; Coccidioides.
 REFERENCE 1 (bases 1 to 978)
 AUTHORS Gardner, M.J. and Cole, G.T.
 TITLE Analysis of gene expression in Coccidioides posadasii mycelia and
 spherules via expressed sequence tags
 JOURNAL Unpublished (2003)
 COMMENT Other ESTs: EST702337
 CONTACT: Gardner MJ
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301 838 3519
 Fax: 301 838 0208
 Email: gardner@tigr.org.
 FEATURES
 source
 1..978
 Location/Qualifiers
 /organism="Coccidioides posadasii"
 /mol_type="mRNA"
 /strain="C735"
 /db_xref="taxon:199306"
 /clone="CIDAU34"
 /dev_stage="saprobic phase (mycelia)"
 /lab_host="E. coli DH10B, T1 phage resistant"
 /clone_lib="Coccidioides posadasii saprobic phase cDNA
 library, 2 to 4 kb"
 /note="Vector: pEXpress 1; Site_1: Not I; Site_2: Eco RV;
 Coccidioides posadasii saprobic phase cDNA library, size
 fractionated cDNA 2 to 4 kb"
 ORIGIN
 Query Match 2.3%; Score 47.6; DB 7; Length 978;
 Best Local Similarity 50.0%; Pred. No. 0.46;
 Matches 119; Conservative 0; Mismatches 119; Indels 0; Gaps 0;
 QY 783 CCAATTCCTCCCGCCAGTGTGCTTCAAGTCCAGCGCGGATGTGTGCGGC 842
 DB 732 CTTTCTCATCTCGCATATGCGCGCGCTGTGTGTGCTGTCCATGCGCGCTACT 791
 QY 843 TGTGTGACATATGAGCTTGTGAGATGCTTGTGACATCTTGGCGGGCTATGAGCAG 902
 DB 792 TGTGTGCGGTGTGTCATATGCGCGCGCTGTGTGCGCATATGTCATGCGCGCTGT 851
 QY 903 AGTGTGACGCGGTGTGTCATATGAGTCAATGAGCGGAAATGCCCTCCACCGAGA 962
 DB 852 TGTGTGCGGTGTGTCATATGAGTCAATGAGCGGAAATGCCCTCCACCGAGA 911
 QY 963 CCGTGTAACTATCTCCCTGCAATCTCTCTGTGTGCGCGGTGTGTGCGGGTGTG 1020
 DB 912 CATGCGCGCGGTGTGTGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 969
 RESULT 5

BX905657/c
 LOCUS BX905657 635 bp DNA linear GSS 07-JUN-2004
 DEFINITION Leishmania braziliensis GSS, clone LBAF018C06, genomic survey
 sequence.
 ACCESSION BX905657 GI:40734125
 VERSION BX905657.1
 KEYWORDS GSS; genomic survey sequence.
 SOURCE Leishmania braziliensis
 ORGANISM Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
 Leishmania; Leishmania braziliensis species complex.
 REFERENCE 1
 AUTHORS Laurentino, E.C., Ruiz, J.C. and Cruz, A.K.
 TITLE GSS analysis of the Leishmania braziliensis genome
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 635)
 AUTHORS Cruz, A.K.
 TITLE Direct Submission
 JOURNAL Submitted (07-JAN-2004) Cruz A.K., University of Sao Paulo,
 Department of Molecular and Cell Biology, FMRP, Avenida
 Bandeirantes 3900, Ribeirao Preto, SP, 14049-900, BRAZIL
 COMMENT Clone requests: akcruz@fmrp.usp.br.
 FEATURES
 source
 1..635
 Location/Qualifiers
 /organism="Leishmania braziliensis"
 /mol_type="genomic DNA"
 /strain="MHOM/BR/75/M2904"
 /db_xref="taxon:5660"
 /clone="LBAF018C06"
 ORIGIN
 Query Match 2.3%; Score 47; DB 9; Length 635;
 Best Local Similarity 48.7%; Pred. No. 0.61;
 Matches 128; Conservative 0; Mismatches 135; Indels 0; Gaps 0;
 QY 780 CCGCCCACTGCTCCCGCCAGTGTGCTTCAAGTCCAGCGCGGATGTGTGTC 839
 DB 382 CGCGAGAGTCCAGTGTGCGCTGTGTGAGCGCAGAGGTGCACTGCGCGCTGC 323
 QY 840 GCGTGTGACATATGAGCTTGTGAGATGCTTGTGACATCTTGGCGGGCTATGAGC 899
 DB 322 TCCCTGAGAGATCCATGATCTTCTGTGATATCTGATATCTGATATCTGATATCTG 263
 QY 900 AGAGTGTGAGCGCGGTGTGTCATATGAGTCAATGAGCGGAAATGCCCTCCACCGA 959
 DB 262 AGAGTGTGAGCGGTGTGTCATATGAGTCAATGAGCGGAAATGCCCTCCACCGA 203
 QY 960 GGACCTGTAACTATCTCCCTGCAATCTCTCTGTGTGCGCGGTGTGTGCGGGTGTG 1019
 DB 202 GCGCGGTATCATATGCGCGGTATCATATGCGCGGTATCATATGCGCGGTATCAT 143
 QY 1020 GTGCGAGCGATATGCGGTGCGC 1042
 DB 142 GTTGTGCGCGGTATGCGGTGCGC 120
 RESULT 6
 LBAF018C06/c
 LOCUS LBAF018C06 635 bp DNA linear GSS 23-JUN-2003
 DEFINITION Leishmania braziliensis GSS, clone LBAF018C06, genomic survey
 sequence.
 ACCESSION BX545484 GI:32169590
 VERSION BX545484.1
 KEYWORDS GSS; genomic survey sequence.
 SOURCE Leishmania braziliensis
 ORGANISM Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
 Leishmania; Leishmania braziliensis species complex.
 REFERENCE 1
 AUTHORS Laurentino, E.C., Ruiz, J.C. and Cruz, A.K.
 TITLE GSS analysis of the Leishmania braziliensis genome
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 635)

AUTHORS Cruz, A.K.
TITLE Direct Submission
JOURNAL Submitted (17-JUN-2003) Cruz A.K., University of Sao Paulo,
Department of Molecular and Cell Biology, FMRP, Avenida
Bandeira 3900, Ribeirao Preto, SP, 14049-900, BRAZIL

COMMENT Clone requests: akcruz@fmrp.usp.br.
FEATURES Location/Qualifiers

source

1. 635
/organism="Leishmania braziliensis"
/mol_type="genomic DNA"
/strain="MHOM/BR/75/M2904"
/db_xref="taxon:5660"
/clone="UBA018C06"

ORIGIN

Query Match 2.3%; Score 47; DB 9; Length 635;
Best Local Similarity 48.7%; Pred. No. 0.61;
Matches 128; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 780 GCCCAACTCGCTCCGCCAGTGTCTTCACTTTCGTAGGCGCCGCACTTGTGTGC 839
DB CGCGAGCTCCACGTTGCCGTTGTGACGCGAGAGGTGCCACGCGCTGTGC 323
QY 840 GCGTGTGGCAGCATAGGCTTGGGAAGTGTCTTGTGACATCTTGGCGGCTATGAGC 899
DB 322 TGCTCAGAGAGATCCATGACTTCTCTCGTAGATCTGTGATCTCCACCTGACGCGGA 263
QY 900 AGAGTGGCAGGCGGCTCGTGCCCTTAAAGTCAATGAGCGGAATGCCCTCCACGGA 959
DB 262 AGAGTGTCTGTCTGTGCTACGCTTGTGTGAGCTCCGCGAACAATCCCGACAGA 203
QY 960 GACCTGTGTAATTAATCTTCCCTGCAATCTCTCTCTGTGTGCGCTGTGTGGGGTGT 1019
DB 202 GCGCGGTACATAGCCCTGAAGACCTCTCGCTCCCATGTAGAGAGCCTTGTGTGCT 143
QY 1020 GTGCGCAGCATCTGTCTGTGC 1042
DB 142 GTTGGCGCGCTGTGAGACTGC 120

RESULT 7

CO015154/c 855 bp mRNA linear EST 09-JUN-2004
LOCUS EST785536 Coccidioides posadasii saprobic phase cDNA library,
DEFINITION greater than 4kb Coccidioides posadasii cDNA clone CIBG75 5' end,
mRNA sequence.

ACCESSION CO015154
VERSION CO015154.1 GI:48522043
KEYWORDS EST.

SOURCE Coccidioides posadasii
ORGANISM Coccidioides posadasii
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;

REFERENCE 1 (bases 1 to 855)
Gardner, M.J. and Cole, G.T.

AUTHORS Analysis of gene expression in Coccidioides posadasii mycelia and
TITLE spherules via expressed sequence tags

JOURNAL Unpublished (2003)

COMMENT Other_ESTS: EST785535

CONTACT: Gardner MJ
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA

TELE: 301 838 3519
FAX: 301 838 0208

EMAIL: gardner@tigr.org
Seq primer: M13 Reverse.

FEATURES Location/Qualifiers

1. 855

/organism="Coccidioides posadasii"
/mol_type="mRNA"
/strain="C735"
/db_xref="taxon:199306"
/clone="CIBG75"

ORIGIN

Query Match 2.3%; Score 47; DB 7; Length 855;
Best Local Similarity 52.9%; Pred. No. 0.65;
Matches 101; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 838 CGCGCTGTGGCAGCATAGGCTTGGGAAGTGTCTTGTGACATTTGGCGGCTATGGA 897
DB 820 GTGCTTGTGGCGCTGTGCTTCATGCGCCCGCTTGTGGCATATGTCATGCGCCG 761
QY 898 GCAGAGTGGCAGGCGCGCTGTGCGCTTAAAGTCAATGAGCGGGAATGCCCTCCACC 957
DB 760 GTGCTTGTGGCGCTGTGCTGTGAGCGCTGTGCTTCATGAGCGCGCTGTGTGACCATC 701
QY 958 GAGACCTGTGTAATTAATCTTCCCTGCAATCTCTCTCTGTGTGCGCTGTGTGGGGTCT 1017
DB 700 ATGTCCATGAGCGCGCGCTGTGCTTGTGSCCATCATGTCCATGAGCGCGCTGTGTGCGCGTTC 641
QY 1018 GTTGGCGGAGC 1028
DB 640 GTGTCCATGAGC 630

RESULT 8

CF823313/c 909 bp mRNA linear EST 01-APR-2004
LOCUS EST700695 Coccidioides posadasii saprobic phase cDNA library, 2 to
DEFINITION 4 kb Coccidioides posadasii cDNA clone CIDAK7 5' end, mRNA
sequence.

ACCESSION CF823313
VERSION CF823313.1 GI:45929370
KEYWORDS EST.

SOURCE Coccidioides posadasii
ORGANISM Coccidioides posadasii
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;

REFERENCE 1 (bases 1 to 909)
Gardner, M.J. and Cole, G.T.

AUTHORS Analysis of gene expression in Coccidioides posadasii mycelia and
TITLE spherules via expressed sequence tags

JOURNAL Unpublished (2003)

COMMENT Other_ESTS: EST700694

CONTACT: Gardner MJ
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA

TELE: 301 838 3519
FAX: 301 838 0208

EMAIL: gardner@tigr.org
Seq primer: M13 Reverse.

FEATURES Location/Qualifiers

source

1. 909

/organism="Coccidioides posadasii"
/mol_type="mRNA"
/strain="C735"
/db_xref="taxon:199306"
/clone="CIDAK77"
/dev_stage="saprobic phase (mycelia)"
/lab_host="E. coli DH10B, T1 phage resistant"
/clone_lib="Coccidioides posadasii saprobic phase cDNA
library, 2 to 4 kb"
/note="Vector: pExpress 1; Site 1: Not I; Site 2: Eco RV;
Coccidioides posadasii saprobic phase cDNA library, size
fractionated cDNA 2 to 4 kb"

ORIGIN

Query Match 2.3%; Score 47; DB 7; Length 909;

[illegible]

| ORGANISM | REFERENCE | AUTHORS | TITLE | JOURNAL | COMMENT |
|--|--------------------|--|-------------------|-------------------------|---|
| Drosophila melanogaster | 1 (bases 1 to 925) | Genoscope. | Direct Submission | Submitted (02-JUN-1999) | Genoscope (E-mail : segre@genoscope.cns.fr BP 191 91006 EVRY cedex - FRANCE) |
| Enkaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. | | | | | |
| FEATURES | | | | | |
| Source | | | | | |
| 1..925 | | | | | |
| Location/Qualifiers | | | | | |
| /organism="Drosophila melanogaster" | | | | | |
| /mol_type="genomic DNA" | | | | | |
| /db_xref="taxon:7227" | | | | | |
| /clone="BACR19D1c" | | | | | |
| /clone_1lb="RPCT-98" | | | | | |
| /note="end : TET3" | | | | | |
| ORIGIN | | | | | |
| Query Match | | | | | |
| Best Local Similarity | 2.3%; | Score 46.6; | DB 9; | Length 925; | |
| Matches | 49; | Conservative | 152; | Mismatches 150; | Indels 0; |
| | | | | Gaps 0; | |
| QY | 725 | GCCGCGTACCAACCAATTACCTCTGCTTAACATCTGGGGAGATGGTACCGCC | 784 | | |
| DB | 920 | SCSCSCSSBSCSSSMSTSSBSBSCSSBSSSTSSMSBSSBSSSSSSSSTSS | 861 | | |
| QY | 785 | AACCTGCTCCCCCAGAGTGTCTTACACTTCGTGAGCGCCGCGATGCTGTGTGCGGTG | 844 | | |
| DB | 860 | SACVKNASSSCGCGCGGMBCCCMCSCSSSSCCGSASARGVAVRASGAGKRGSSGGGSA | 801 | | |
| QY | 845 | TTGCGACATGAGCGCTTGGGAAGTGTCTTGACATCTTGACCGAGCTATGACAGAGG | 904 | | |
| DB | 800 | SHSSSSACBSSSSSCSACMSWASSSSSASRSRSGGAGSGASASRSSSSSSSASASG | 741 | | |
| QY | 905 | TGCGAGCGCGCTGTGTGCTTTAAGTCATGACGGCGGAATGCCCTTCACCGAGACG | 964 | | |
| DB | 740 | VVSSASSSSSCSSSVSCSSVASSMSCSBSSSASASSSSSSSASACSCCCTTSMSC | 681 | | |
| QY | 965 | TGTTAATTAATCTCCCTGACATCTCTCTCTGTGTGCGGTGCTGTGTGTGCGGTG | 1024 | | |
| DB | 680 | SCSTASMSARSSSSSSSSSCSSSSMSASASSSSSSSSSSSSSSSSGACBSMSWG | 621 | | |
| QY | 1025 | CAGGATATCTGCGGACGCTGAGTCCAGGGAGGGAGCTGTGCAGTGA | 1075 | | |
| DB | 620 | GGSGSVASMSGSSSVSSSGGSGSGGGGAGGAGGSSSSSSGSGSGSV | 570 | | |
| RESULT 13 | | | | | |
| CNS01213 | | | | | |
| LOCUS | | | | | |
| DEFINITION | | | | | |
| ACCESSION | | | | | |
| VERSION | | | | | |
| KEYWORDS | | | | | |

| SOURCE ORGANISM | REFERENCE AUTHORS | JOURNAL TITLE | COMMENT | FEATURES | ORIGIN |
|-------------------------------------|--|---------------|--|-------------------------------------|--------|
| Drosophila melanogaster (fruit fly) | Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. | Genoscope. | Direct Submission | Location/Qualifiers | |
| | 1 (bases 1 to 645) | | Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr - Web : www.genoscope.cns.fr) | 1. 645 | |
| | | | determination of this BAC-end and sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk/. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billard at CBPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11. | /organism="Drosophila melanogaster" | |
| | | | | /mol_type="genomic DNA" | |
| | | | | /db_xref="taxon:7227" | |
| | | | | /clone="BACN08C07" | |
| | | | | /clone_11b="DrosBAC" | |
| | | | | /plasmid="pBeloBAC11" | |
| | | | | /note="end : 17" | |
| Query Match | 2.3%; Score 46.2; DB 9; Length 645; | | | | |
| Best Local Similarity | 12.0%; Pred. No. 0.99; | | | | |
| Matches | 55; Conservative 127; Mismatches 276; Indels 1; Gaps 1; | | | | |
| 800 | GTGCTGCTTACGACTTTCGTAGACCGCGGACATTCGTGTCGGCGCTGTGACAGATPAGCC | 859 | | | |
| 129 | KKGGGAGNNNGTNGTGTGTTGNNNGGGGGGNNNNNGNNNTTNTTNGGTTTNNNGNNN | 188 | | | |
| 860 | TTGGGAAGTCTGTGTGACATCTTGCGCGGCGCTATGAGCAGATGCGAGCGCGCTCG | 919 | | | |
| 189 | NNSCCSCTCNNNNTTNTTSSSSSSSNNNTNNNNNNNNNNNNSSSSSSSSNNN | 248 | | | |
| 920 | TGCGCTTTAAGTATGAGCGGCGCAATGCGCTTCCACGAGACCTGTTAATTACTTCC | 979 | | | |
| 249 | NSGGSSSSNNSSNNSSGSSSSNSTSSNNNNNTTNNNNSSNNNNSSSTNSNNSSSS | 308 | | | |
| 980 | CTGCATCTCTCTCTCGTGCCCTCGTGTCTGCGGGTGTGTGCGCACCATCTGCTCTC | 1039 | | | |
| 309 | TSTTNTSNTSSNNSSATSSSSSSSGTTTSSSSSSSSSSSSNNNNNNNNNNNNSS | 368 | | | |
| 1040 | GGCAGTGTGCTCAGAGGAGGAGGCGTGTGACAGTGAACCGCGCTGATAGCGTTCGCT | 1099 | | | |
| 369 | NNSNSATTTST-NSGSSSSSSSSSSSSSTSTSTSTSTSSNNSSSTTSSSTTNTST | 427 | | | |
| 1100 | CGCGGGTAAACATGTTTCCCCACGACACTATGTGCGAGAGCGACGCCGACGACGTG | 1159 | | | |
| 428 | SSSSSSNSTGSSSSSTTTTNSASNNNNNNNNNNSSSSNAANSTSTTSSNTTSSNNNS | 487 | | | |
| 1160 | TCACATCAGATCTCTCCGACCTTACTATCACCACCACTGTGAAGAGCTCCACCACTGGA | 1219 | | | |
| 488 | SCNTTTSANTSTSTNTTNTAGNNNNATSSSSNNNNNNNNNNSSGSGTSGNTSSSSC | 547 | | | |
| 1220 | TTAACGAGAGTGTCTCAGCGCCCTGCTCGGCTCGTGAC | 1258 | | | |
| 548 | TGTBSGNSNSNNSSNNSSCTGNSNSSTTTGSS | 586 | | | |

| Accession | Query Match | Score | DB | Length |
|--|-----------------------|-------|-----------------|----------------------------------|
| CP821532 | Query Match | 2.2% | Score 45.8 | DB 7; Length 881; |
| CP821532.1 | Best Local Similarity | 55.3% | Pred. No. 1.3; | |
| GI:45927589 | Matches | 89; | Conservative 0; | Mismatches 72; Indels 0; Gaps 0; |
| EST. | | | | |
| Coccidioides posadasii | | | | |
| Coccidioides posadasii | | | | |
| Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; | | | | |
| Omygenales; microspor Omygenales; Coccidioides. | | | | |
| 1 (bases 1 to 881) | | | | |
| Gardner, M.J. and Cole, G.T. | | | | |
| Analysis of gene expression in Coccidioides posadasii mycelia and | | | | |
| spherules via expressed sequence tags | | | | |
| Unpublished (2003) | | | | |
| Other ESTs: EST568913 | | | | |
| Contact: Gardner M.J. | | | | |
| The Institute for Genomic Research | | | | |
| 9712 Medical Center Drive, Rockville, MD 20850, USA | | | | |
| Tel: 301 838 3519 | | | | |
| Fax: 301 838 0208 | | | | |
| Email: gardner@igr.org | | | | |
| Seq primer: M13 Reverse. | | | | |
| Location/Qualifiers | | | | |
| 1..881 | | | | |
| /organism="Coccidioides posadasii" | | | | |
| /mol_type="mRNA" | | | | |
| /strain="C735" | | | | |
| /db_xref="taxon:199306" | | | | |
| /clone="CID465" | | | | |
| /dev_stage="saprobic phase (mycelia)" | | | | |
| /lab_host="E. coli DH10B, T1 phage resistant" | | | | |
| /clone_lib="Coccidioides posadasii saprobic phase cDNA | | | | |
| library, 2 to 4 kb" | | | | |
| /note="Vector: pExpress 1; Site_1: Not I; Site_2: Eco RV; | | | | |
| Coccidioides posadasii saprobic phase cDNA library, size | | | | |
| fractionated cDNA 2 to 4 kb" | | | | |
| ORIGIN | | | | |
| Query Match | | | | |
| Best Local Similarity | | | | |
| Matches | | | | |
| Conservative | | | | |
| Mismatches | | | | |
| Indels | | | | |
| Gaps | | | | |
| DB | | | | |
| QY | | | | |
| CG855542/c | | | | |
| LOCUS | | | | |
| DEFINITION | | | | |
| ACCESSION | | | | |
| VERSION | | | | |
| KEYWORDS | | | | |
| ORGANISM | | | | |
| SOURCE | | | | |
| ORGANISM | | | | |
| Zeae maye | | | | |
| Zeae maye | | | | |
| Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; | | | | |
| Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD | | | | |
| Clade; Panicoideae; Andropogoneae; Zea. | | | | |
| 1 (bases 1 to 753) | | | | |
| Bharti, A.K., Young, S., Kachok, S., Keiser, G., Bronzino, A.C., | | | | |
| Zohovetz, V., Fuks, G., Yu, Y., Wing, R. and Messing, J. | | | | |
| Sequencing of the maize genome at pEIR (2003c) | | | | |
| Unpublished (2003) | | | | |
| Contact: Bharti, A.K. | | | | |
| Dr. Joachim Messing's lab | | | | |

The Plant Genome Initiative at Rutgers, Waksman Institute, Rutgers University
190 Frelinghuysen Road, Piscataway, NJ 08854, USA
Tel: 732 445 3801
Fax: 732 445 5735

Email: bharti@waksman.rutgers.edu
Seq primer: SP6

Class: BAC ends

High quality sequence start: 394.

FEATURES

source

location/Qualifiers
1..753
/organism="Zea mays"
/mol_type="genomic DNA"
/cultivar="B73"
/db_xref="taxon:4577"
/clone="ZMMBc0257G23"
/lab_host="E. coli DH10B"
/clone_lib="ZMMBc (EcoRI)"
/note="vector: pIARBAC2.1; Site_1: EcoRI; Site_2: EcoRI"

ORIGIN

```

Query Match          2.2%  Score 45.4; DB 9; Length 753;
Best Local Similarity 54.0%; Pred No. 1.7;
Matches 116; Conservative 0; Mismatches 96; Indels 3; Gaps 1;

QY 775 GTAGCCGCCCAACTGCTCCCCAGTGTGCTTCACTTTGTAAGCGCCGGCATTGCT 834
    |||||
DB 617 GTGCGCGCCGCGAGCTGCTGCTTCACTGTGCTGCTGCGCCCGCCGCGCATGTC 558
    |||||

QY 835 GGTGC--GGCTGTGGCAGCATAGGCTTGGGAAGTGTGTGACATCTTGGCGGCG 891
    |||||
DB 557 GTTTCCTTCGCGCTATGTCATGCGCCCTGCTGTGTCGACATGCTGCTGTC 498
    |||||

QY 892 TATGAGCAGAGAGTGCAGCGCGCTCGTGCCTTTAAGTCAATGAGCGCGGAATGCC 951
    |||||
DB 497 ACTGTGCTTCATGCGCGCGCTCGTATGCTGCTGCTCCATGAGTGCAGAGATCC 438
    |||||

QY 952 TCCACCGAGGACTGTAACTTACTCCCTGCCAT 986
    |||||
DB 437 TCCACCTAATTAAGTAGGCTCTCCGAGTCAAT 403
    |||||

```

Search completed: February 25, 2005, 07:35:09
Job time : 6010 secs